UBLICATIONS DE L'INSTITUT DE STATISTIQUE
DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS,
LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

Benoît MANDELBROT

Contribution à la Théorie Mathématique des JEUX DE COMMUNICATION

M. P. SCHÜTZENBERGER

REMARQUES SUR LE PROBLÈME DU CODAGE BINAIRE

VOL. II - FASCICULES 1-2 - 1953

PARIS
INSTITUT HENRI POINCARÉ
11, Rue Pierre Curie

Toute la correspondance relative aux publications
doit être envoyée à l'adresse :

INSTITUT DE STATISTIQUE DE L'UNIVERSITE DE PARIS

INSTITUT HENRI POINCARÉ - 11, Rue Pierre Curie - PARIS (Ve)

Les manuscrits doivent être envoyés à *M. Daniel DUGUÉ*,
à l'adresse précédente.

Abonnements : Pour la France 1.200 francs français
Pour l'Etranger 1.500 francs

Vente au numéro : *(fascicule de 50 pages environ)*
Pour la France 350 francs français
Pour l'Etranger 400 francs

UBLICATIONS DE L'INSTITUT DE STATISTIQUE DE L'UNIVERSITÉ DE PARIS

MÉMOIRES ET CONFÉRENCES SUR LE CALCUL DES PROBABILITÉS, LA STATISTIQUE THÉORIQUE ET APPLIQUÉE, L'ÉCONOMÉTRIE

VOL. II

PARIS

INSTITUT HENRI POINCARÉ

11, Rue Pierre Curie



PUBLICATIONS DE L'INSTITUT DE STATISTIQUE DE L'UNIVERSITÉ DE PARIS
Volume 2 FASCICULES 1 et 2 1953

CONTRIBUTION A LA THÉORIE MATHÉMATIQUE DES

JEUX DE COMMUNICATION

par

BENOIT MANDELBROT
Ancien Elève de l'Ecole Polytechnique, Docteur es-Sciences
Ingénieur aux Laboratoires d'Electronique et de Physique Appliquées

THÈSE SOUTENUE LE 16 DÉCEMBRE 1952

ARGUMENT

LES JEUX DE STRATÉGIE DE LA COMMUNICATION EN TANT QUE MODÈLES PHYSIQUES

Les jeux de stratégie, au sens mathématique, ainsi que le critère de "minimax de gain", ont été introduits par Borel et von Neumann pour fournir des modèles au comportement économique. Wald les a utilisés, avec les critères de minimax et de Bayes, pour obtenir des modèles du comportement inductif en statistique.

On fait remarquer que l'ensemble des problèmes de la communication peut, lui aussi, être étudié dans le cadre unifié d'une nouvelle classe de "jeux de communication" :
les jeux à trois joueurs :

émetteur - nature - récepteur

et ceux obtenus comme séquence de tels jeux à trois joueurs.

Dès que le nombre de joueurs dépasse deux, on introduit la possibilité de coalition entre joueurs. En fait tous les jeux de communication consistent à s'arranger de façon que la transmission déforme le message le moins possible : ce fait peut se traduire en une coalition entre les joueurs ultimes contre la nature.

Le cadre stratégique peut être utilisé directement en vue des problèmes techniques de recherche des meilleures méthodes de transmission, étant données les propriétés statistiques du message et les propriétés physiques de la Nature, supposées toutes

connues : les travaux de C.E. Shannon se rangent à posteriori dans cette catégorie.

Mais on peut également utiliser ce cadre inversement. A cet effet, on construit des jeux parfaits, en des sens déterminés, et on montre que les résultats observables que l'on en déduit sont effectivement vérifiés par certains phénomènes physiques.

Deux exemples sont étudiés en détail. D'une part les lois de la thermodynamique peuvent être considérées comme constituant la description de la nature, d'intérêt pour un jeu optimal minimax émetteur-nature. Par ailleurs, la structure statistique de la langue est celle-là même qui permet une certaine coalition minimin entre émetteur et récepteur.

Les jeux de communication peuvent ainsi jouer le rôle de modèles abstraits pour la physique.Le procédé d'induction que nécessitent ces modèles est étudié en détail.

INTRODUCTION

0. 1 - GÉNÉRALITÉS SUR LA THÉORIE DES COMMUNICATIONS

0.1.1 - DÉFINITION

Nous appelons Théorie des Communications l'ensemble des concepts mathématiques, lois physiques et procédés techniques, relatifs à une certaine classe de jeux de stratégie à trois joueurs :

1. ÉMETTEUR - 2. NATURE - 3. RÉCEPTEUR

Dans la plupart des cas, un jeu de communication est suivi d'un autre jeu où les rôles de l'émetteur et du récepteur sont intervertis. Il arrive même (M. R. Fortet a fait remarquer que c'est le cas en Radar) qu'il soit nécessaire de considérer globalement deux tels jeux successifs. Cependant, tel ne sera pas le cas dans les exemples étudiés dans ce mémoire : émetteur et récepteur seront deux interlocuteurs symétriques, coalisés contre la Nature, et il ne sera donc pas nécessaire de considérer le jeu (3. 2. 1) pour décrire le jeu (1.2.3) (sauf si l'on veut étudier le mécanisme par lequel la coalition entre 1 et 2 a pu s'établir).

Précisons la nature des stratégies du jeu (1.2.3). D'après Kolmogoroff (1941) et Wiener (1948), le message est un processus stochastique. Donc le fait de le produire et de l'encoder constitue pour l'émetteur une "stratégie mixte". Le récepteur, ignorant le message, mais connaissant le code, joue une "stratégie pure" pour décoder, c'est-à-dire identifier le message reçu.

Pour la Nature, la règle du jeu est de perturber les messages d'une certaine façon, connue statistiquement, comme conséquence des lois de la physique, à la fois de l'émetteur et du récepteur, c'est ce qui permet à ces joueurs de se coaliser contre la Nature.

Les jeux de communication sont donc à priori très spéciaux, et en particulier, très asymétriques. A part cela, ils peuvent différer très fortement les uns des autres, suivant la nature des messages dont il s'agit. L'existence d'une théorie des communications unique repose donc sur un postulat spécial, qui uni-

fie, sur le plan fonctionnel, des phénomènes physiques, biologiques et humains :

Nous admettons que les conditions fonctionnelles des processus de communication (et de contrôle) sont susceptibles d'une étude abstraite suivant les normes de rigueur de la physique, quantitative et conceptuellement homogène, et de plus indépendante de leurs réalisations physiques, dont applicable aussi bien aux organismes vivants qu'aux mécanismes artificiels.

M. Wiener, à qui l'on doit l'exposé le plus frappant de ce principe (1948) a proposé d'appeler "Cybernétique" la science qui en étudierait les applications.

0.1.2 - OBJET ET PLAN DU MÉMOIRE

Un aspect purement descriptif de la science des communications consisterait, dans chaque cas particulier, à compléter la description générale du § précédent, par des détails aussi précis que possible.

Il est toutefois évident que, d'un tel point de vue, il était inutile de se placer d'emblée dans un cadre aussi général que celui de la théorie des jeux de stratégie (que nous appelons aussi "stratégique"). On cherchera donc à remplacer la description exhaustive par une description plus globale, justifiant l'utilisation des jeux. Ceci se fera en deux étapes. La première étape cherchera à définir la "valeur" d'un jeu de communication donné. La deuxième étape cherchera à expliquer cette "valeur" en la déduisant d'une fonction de gain et d'un critère stratégique (Si une telle explication se révèle impossible, le jeu n'étant pas optimal, ce procédé permet quand même de définir le rendement d'un jeu de communication donné).

Chacune des deux parties du présent mémoire sera consacrée à l'une de ces étapes. Etant données les difficultés de toute étude abordant directement des jeux à trois joueurs, nous remplacerons toujours ceux-ci par des groupes de trois jeux à deux joueurs, obtenus en faisant des hypothèses simplificatrices sur chacun des trois joueurs successivement.

La première partie comportera donc des compléments à la théorie des jeux à deux joueurs : elle définira la famille d'où doivent être tirées les fonctions de gain de la communication : c'est la famille des informations temporelles. On verra comment on y rattache les concepts de durée et de Démon de Maxwell.

Quant aux trois exemples de la deuxième partie, ils nous montreront que la Thermodynamique telle qu'elle a été bâtie à partir d'expériences, et la structure statistique des langues, qui résulte des décomptes empiriques, correspondent respectivement à ce qui résulterait d'un conflit minimax entre Emetteur et Nature, et d'une certaine coalition optimale entre Emetteur et Récepteur. La fonction à laquelle s'appliquent ces deux critères est dans les deux cas déduite de l'"information sélective" : la première fois c'est une fonction de gain proprement dit, moins une fonction de risque; dans le deuxième cas, gain moins coût. Ceci montre que, parmi les données de base, nécessaires au développement ultérieur de la théorie, deux au moins ne sont pas des résultats autonomes, indépendants des bases de cette théorie.

Ces exemples ont tous deux un caractère "physique", dans un sens hétérodoxe pour le deuxième, mais finalement peu discutable, de ce terme. Ils ont été choisis en partie pour la valeur du support qu'ils offraient à l'intuition pour faire évoluer les concepts fondamentaux et les rendre mieux aptes à attaquer des problèmes réels. Ceux-ci forment l'objet premier de l'étude théorique des moyens propres à améliorer l'utilisation du temps et de l'énergie dans les méthodes de transmission, que nous poursuivons aux Laboratoires d'Electronique et de Physique Appliquées.

Dans tout ce mémoire, on suppose connus les résultats du mémoire classique de Shannon (1948), qui ne sont quelquefois redémontrés que pour rendre plus claire la suite des raisonnements.

Le travail est organisé de telle façon que l'on peut court-circuiter les considérations thermodynamiques en omettant les chapitres 4 et 5. De même, si l'on ne s'intéresse qu'aux applications nouvelles du codage séquentiel et pas aux fondements, on peut ne lire que les chapitres 6 et 7 (qui ont d'ailleurs été rédigés avant les chapitres 1 à 5).

La théorie statistique de la langue a déjà été exposée dans Mandelbrot (1951, a, b) et Mandelbrot (1952, c); la théorie des informations et durées et des démons de Maxwell dans Mandelbrot (1952, a, b). (Ces deux Notes sont toutefois assez imprécises, et ont été modifiées légèrement dans l'exposé complet qui suit).

0.1.3 - PROBLÈMES PRATIQUES DE COMMUNICATION

Il est bon de préciser ici la relation entre la Théorie des Communications et les problèmes pratiques de transmission.

On peut dire de façon tout à fait générale que ceux-ci ont pour but, après une suite d'opérations où le message est déformé de façon arbitraire, de reconstituer un objet qui, du point de vue de l'organe récepteur, soit "équivalent" au message initial.

On peut dire aussi qu'il s'agit de construire un cycle fermé relativement à cette équivalence, en adaptant certaines phases du cycle aux moyens de transmission fixés, et en adaptant alors les moyens de transmission non fixés aux phases du cycle qui restent encore libres.

On peut dire encore qu'il s'agit d'établir l'adaptation statistique entre :

- d'une part, celles des propriétés d'un message qui justifient le désir que l'on a de le transmettre,
- d'autre part, celles des propriétés de la ligne de transmission qui conditionnent les possibilités de transmission qu'offre cette ligne.

Pour cela, on s'efforcera de représenter le message et la ligne de transmission par des algorithmes appropriés et de les caractériser chacun par un ou plusieurs paramètres : des "variables d'état". Celles-ci rendent la théorie plus simple, en permettant d'exprimer l'adaptation par des égalités entre nombres, et par conséquent la rendent plus accessible à l'intuition.

Remarquons que, sans que le terme "variable d'état" ait été utilisé, les lignes de transmission continues ont depuis longtemps été ainsi caractérisées par deux grandeurs : le rapport

signal/bruit et la largeur de bande W, et par une fonction de ces grandeurs : la capacité Wlog (1 + S/B). (Nous montrerons que les lignes discrètes possèdent aussi des variables d'état: Cf. § 6. 2.2.).

Par ailleurs, les messages discrets sont très bien caractérisés par leur "information sélective" de Shannon et Wiener. C'est le cas de la langue considérée comme suite de mots ou de lettres. De même, lorsqu'un message quelconque est déjà artificiellement quantifié, c'est la valeur de l'information sélective qui conditionne la meilleure transmission : elle est donc tout à fait priviligiée. Mais cette information dépend de la manière dont le message a été rendu discret : elle peut donc prendre des valeurs très différentes pour deux images "équivalentes" du point de vue du récepteur. On en voit déjà un exemple dans le cas du signal continu du temps continu, où les indéterminations de temps τ et de puissance δ sont liées par $\tau\delta = kT/2$ (§ 4.1). L'information sélective dépend du τ choisi et n'est pas une variable d'état, tandis que l'"information spécifique de Fisher" se trouve égale à $1/\delta\tau$ et caractérise donc le signal en tant que "variable d'état". Cette expression, sans possèder toutes les propriétés de l'information sélective, conserve celles qui justifient intuitivement le terme "information".

Il existe d'autres telles informations (§ 2.2) et on ne résoud donc pas le problème en disant qu'il y a adaptation lorsque toute l'information utile est transmise à meilleur compte : il reste encore à identifier cette information - variable d'état, ce qui est encore très difficile. Il est donc extrêmement inquiétant de constater que, pratiquement, tous les modèles utilisés en théorie des Communications sont linguistiques, car il en résulte qu'ils sont exagérément simples, non seulement du point de vue de la manipulation technique, ce qui serait très heureux, mais aussi du point de vue du rôle que joue l'information sélective, ce qui est dangereux.

0.1.4 - REMARQUE METHODOLOGIQUE

Nous allons parler de Thermodynamique et aussi de Linguistique. Les objets du présent mémoire chevauchent donc par dessus les limites des sciences physiques, biologiques et humaines. Mais ils sont liés entre eux par des thèmes communs : les concepts de stratégie et d'information. L'unité du mémoire est ainsi d'ordre mathématique, mais les calculs proprement dits étant élémentaires, le mémoire ne saurait tirer son intérêt de leur difficulté.

L'unité du mémoire est aussi d'ordre méthodologique. Les problèmes de méthode se sont d'abord posés avec acuité, lorsqu'il a fallu appliquer dans le domaine de la linguistique des procédés qui avaient acquis l'assentiment des chercheurs en Physique classique. Mais il s'est ensuite révélé que même dans des domaines traditionnels, un effort de clarification était très utile. Les problèmes de méthode nous ont donc finalement préoccupé tout le temps, et en particulier au premier chapitre, qui exposera le point de vue stratégique en physique et introduira la distinction entre problèmes directs et inverses, qui nous sera indispensable par la suite.

Les problèmes de méthode montreront l'importance de la position qu'occupent dans la science la théorie des transmissions, qui fournit le concept d'information, et la statistique mathématique, qui fournit le concept de stratégie de comportement inductif. Ces deux disciplines se situent à l'extrême limite de la physique généralement acceptée (et même de la doctrine philosophique de la connaissance qu'utilisent implicitement, et souvent inconsciemment, les physiciens). Il en résulte qu'elles mettent en pleine lumière certaines imprécisions et insuffisances de la physique et que par suite leur importance conceptuelle dépasse en étendue et aussi en profondeur, le champ de leur application directe. Une grande partie de la physique peut s'organiser autour de ces concepts.

L'unité méthodologique de la Théorie des Communications ne se manifeste cependant nullement par une opposition à la Thermodynamique, comme on l'a quelquefois suggéré, puisque, ainsi que nous l'avons annoncé, la Thermodynamique constituera l'étude d'un jeu particulier de notre Théorie.

Un tel rôle unificateur peut à priori paraître surprenant de la part de la Théorie des Jeux de stratégie, étant donné que celle-ci est née de l'acceptation par les mathématiciens du fait que la mathématique classique, qui avait été la base de la physique, ne pouvait fournir de cadre à l'étude de l'économie politique. Au contraire, cette dernière comporte une structure identique à celle des jeux de stratégie (Borel) tels que le poker et les échecs : dans les deux cas, il s'agit d'organiser des mouvements en une stratégie en vue d'un gain. Ceci sembla condamner l'ambition des sciences de la nature; de coordonner et d'unifier, au sein d'une mathématique unique, les descriptions des diverses connaissances humaines, dès que celles-ci deviennent suffisamment élaborées (ambition qui était presque une définition de l'effort des physiciens, et avant eux, des "philosophes"). En l'abandonnant, on crut donc que les sciences humaines nécessiteraient une deuxième mathématique : la mathématique du conflit, et qu'il en résulterait une dualité dans la Science, caractérisée par la dualité du calcul infinitémisal et de la Stratégique.

A côté de la Stratégique, et créée vers la même époque, la théorie de la communication de l'information se posait au contraire, comme nous l'avons vu, en théorie unificatrice.

L'étude présence a été commencée sur la base de la théorie de l'information. Elle a été particulièrement influencée par les réflexions que nous avons faites sur le travail de L. Szilard sur le démon de Maxwell (1929). Mais cette étude nous a rapidement amené à la conclusion que l'information de la "Cybernétique" est seulement une fonctionnelle particulière se rattachant à certaines stratégies,dans des problèmes de comportement inductif,et faisant partie d'une famille très large d'autres informations.

Cette conclusion, qui se base sur les travaux de J. Neyman et A. Wald sur les problèmes de comportement inductif, mène bien loin de l'idée que la stratégique n'est pas un outil approprié à l'étude des disciplines physiques où triomphait le calcul différentiel. Tout au contraire, c'est dans le domaine du physicien,

où les problèmes sont mieux posés, qu'elle montrera le mieux sa puissance d'analyse. Elle finira peut-être même par fournir à la Science l'élément d'unité dont elle était censée pallier l'absence.

Remarque terminologique : Il résultera de ce qui suit que les applications physiques et techniques de la Théorie mathématique du Comportement Inductif sont susceptibles de constituer un corps de doctrine étendu et varié. Il chevauche sur l'ensemble des problèmes qu'on serait tenté de classer dans la "Cybernétique". Cependant, cette dernière n'existe pas encore comme Science indépendante : il y a un nom, une série de présomptions très brillantes, mais de forme souvent non physique, relatives à des caractéristiques fonctionnelles communes à de nombreuses organisations, et enfin des méthodes mathématiques d'analyse, héritées de techniques hétéroclites et par suite suivent mal reliées à ces présomptions.

Dès lors, deux attitudes se présentent à l'esprit, raisonnables l'une et l'autre : la première attitude identifierait la cybernétique à l'ensemble des applications bien étudiées de la théorie du comportement inductif; la deuxième attitude conserverait à la cybernétique son caractère imprécis et provocateur, en lui enlevant tous les problèmes qui ont pris place dans la théorie du comportement, et ne lui laissant que les autres. Nous n'avons pas voulu prendre parti dans cette alternative, et évitons jusqu'à nouvel ordre le terme "cybernétique".

Par ailleurs, nous avons déjà indiqué que nous utiliserons "stratégique" comme substantif pour désigner la "théorie des jeux de stratégie". L'accent sur "jeu" serait maintenant tout-à-fait injustifié. En fait, la stratégique se décomposera en théorie de l'action, dont nous parlerons peu, et théorie de l'expérience, dont nous parlerons beaucoup.

Enfin, il nous semble opportun de discuter la possibilité d'utiliser le terme "sémiologie", tiré par le linguiste F. de Saussure (1916) du grec "semeion" = signe, pour désigner la science des signes et de tout ce qui les concerne, en particulier l'information qu'ils transmettent. "sémiologie" serait donc synonyme de "théorie des communications".Le déplacement de l'accent de l'information au signe est d'ailleurs conforme à leurs importances relatives : l'information n'est qu'un certain nombre rattaché au signe, et l'oeuvre de Shannon un instrument d'étude.

* * *

Le présent mémoire est le résultat de recherches effectuées sous la direction de M. le Professeur Georges Darmois. Je suis particulièrement heureux de lui exprimer ici mon profond respect, et ma reconnaissance pour ses encouragements, ses critiques, et ses conseils. Je le remercie également d'avoir accepté ce travail dans la Collection des Publications de l'Institut de Statistique.

Je prie M. Louis de Broglie, Secrétaire Perpétuel de l'Académie des Sciences, qui m'a fait l'honneur d'accepter la présidence du jury de ma thèse, de trouver ici l'expression de ma respectueuse gratitude.

Je remercie M. Robert Fortet d'avoir bien voulu se joindre au jury de ma thèse.

Ces recherches n'auraient pas pu être poursuivies dans des conditions aussi favorables sans le soutien constant et éclairé de M. le Professeur G.A. Boutry. Je le prie d'agréer l'expression de tout mon respect et de ma plus vive gratitude. Je remercie la Direction des Laboratoires d'Electronique et de Physique Appliquées pour les facilités qu'elle a mises à ma disposition et pour l'autorisation de publier ce mémoire.

Enfin je remercie M. le Professeur D. Gabor, du Collège Impérial de Londres, de son aide inestimable dans la mise au point d'une version antérieure du mémoire.

0.2 - GÉNÉRALITÉS SUR LA PREMIÈRE PARTIE DU MÉMOIRE

0.2.1 - ORIGINE ET MULTIPLICITÉ DES CONCEPTS D'INFORMATION

Le concept d'information joue depuis longtemps dans la Science un important rôle heuristique, que son sens dans la vie courante explique suffisamment. Etant donné que les considérations heuristiques ne sortent pas en général des chapitres de la Science où elles ont pris naissance, il n'y avait pas alors de grave inconvénient à ce que les sens attribués au mot "information" varient de l'un à l'autre de ces chapitres, et la variation était en fait considérable.

L'un des rôles les plus importants était celui qui accompagnait l'interprétation de la loi de Carnot par des échanges ordre-désordre. De ce concept qualitatif se sont depuis dégagés deux concepts quantitatifs, qui malheureusement ne sont nullement interchangeables.

D'une part, l'étude de la transmission de l'"information" a amené les ingénieurs de Télécommunications Hartley (1928) et Shannon (1948) à caractériser les messages par une information, que nous qualifierons de sélective ou séquentielle, et dont la forme fonctionnelle est identique à celle de l'entropie selon L. Boltzmann.

D'autre part, Edgeworth puis surtout R.A. Fisher (1922) ont introduit en statistique un concept de "précision", fonctionnellement, et tout au moins en apparence, très différent du précédent, mais dont certaines propriétés étaient celles de l'entropie (1935, p.47).

Les statisticiens et surtout les ingénieurs des Télécommunications ont fait une abondante publicité aux concepts qui leur servent. Il en résulte que, suivant que l'on a l'une ou l'autre formation, on peut être tenté d'attribuer des propriétés, qui ne sont vraies que pour le concept auquel on est habitué, à d'autres concepts d'information, que l'on est susceptible de rencontrer, et pour lesquels ces propriétés ne sont plus vraies.

Dès lors, l'inocuité relative du mot "information" devient un souvenir du passé, et il ne faut l'employer qu'avec la plus grande prudence.

Cependant, la confusion, potentiellement très grave, reste très répandue : en particulier, on généralise abusivement l'usa-

ge de l'information sélective, qui, bien qu'elle soit, comme nous le verrons, la meilleure dans un certain sens, n'est qu'un cas particulier d'une famille assez large.

N. Wiener n'est d'ailleurs pas étranger à cette confusion, car il a affirmé (1948, p. 76) que sa définition remplaçait celle de Fisher (ce qui n'a pas été confirmé). De même, Shannon (1948, Appendice 2) a laissé croire que l'"information sélective" était la seule à satisfaire aux "axiomes raisonnables" de l'information. Elle devrait donc englober l'information fishérienne, ce qui n'a pas non plus été confirmé, mais avait semblé justifier l'utilisation de l'information sélective dès que les axiomes de Shannon semblaient satisfaits. Par exemple, certains auteurs se contentent de remarquer qu'il "semble intuitif que l'objet de toute mesure est de donner de l'information" et en concluent qu'il s'agit d'information sélective. Or, il se trouve qu'en général il s'agit là d'information fishérienne. Mesure et transmission sont bien deux aspects d'une même réalité, mais seulement à un niveau d'abstraction plus élevé.

De même, de très nombreux travaux s'autorisent de l'origine entropique de l'information sélective pour identifier ces deux notions. Citons Brillouin (1949, 1950). Dans tous les cas, une "interprétation" informationnelle est superposée à posteriori à une thermodynamique supposée déjà réalisée. Par suite, on n'a aucune occasion de montrer le rôle de l'information, ni en particulier de voir pourquoi il s'agit d'information sélective, dans la mesure où l'entropie lui est effectivement analogue ou identique.

Réagissant contre ces conclusions hâtives, la première partie du mémoire comportera en particulier l'examen du problème suivant : A quoi correspond l'usage de plus en plus fréquent du terme "Information" dans les domaines les plus divers de la Science : est-il dû à l'imprécision du terme, qui permet toutes sortes d'interprétations; ou introduit-il entre ces divers domaines un lien réel, dont l'étude pourrait contribuer au progrès de la Physique et de la Technique.

Le seul moyen de remédier à l'état de choses actuel, fort peu satisfaisant, est de recenser tous les concepts acceptables du mot, et de s'interdire d'employer les uns à la place des autres; on ajouterait des qualificatifs pour plus de sûreté.

Pour classer ensuite les divers concepts, il faudra un principe de classement, qui pourrait :

- ou bien résulter à posteriori des régularités observées entre tous les concepts qui ont été réunis, comme l'a fait par exemple D.M. Mackay (1950) dans un travail qui a beaucoup influencé les débuts de celui-ci, mais dont nous avons finalement très peu gardé, car sa classification est très partielle et les propriétés qu'elle utilise, superficielles;

- ou bien être choisi à priori;

- ou bien être dégagé de la comparaison de deux seulement des concepts d'information : sélective et de Fisher, et ensuite étendu à la classification des autres concepts. Nous montrerons qu'un tel principe, qui est dû à M. P. Schutzenberger (1951), loin d'être une simple habileté formelle, a une signification

intrinsèque, et qu'on peut beaucoup tirer de son développement. Ce principe conduit à la définition de l'information temporelle qui englobe les gains admissibles pour les jeux de communication (Toutefois, cette définition n'englobe pas toutes les "informations". On peut par exemple citer les "informations généralisées" de M. Féron, ainsi que les concepts "structuraux" de l'information de Gabor et Mackay (1950). Enfin, l'"information métrique" de Mackay a besoin d'être modifiée pour entrer dans ce schéma).

Pour établir cette signification, il faut des considérations assez abstraites, mais il nous semble précisément que l'échec du concept d'information devant les plus importants des problèmes technologiques concrets est dû au fait que son étude n'était pas suffisamment abstraite et n'était pas allée suffisamment au foud des propriétés de l'information. Nous allons donc tenter cet effort d'abstraction et espérons montrer qû'il permet une théorie unifiée et puissante ("It is no paradox to say that in our most theoretical moods, we may be nearest to our most practical applications").

0.2.2 - DURÉES ET DÉMONS DE MAXWELL

La généralisation de l'information, que nous donnerons d'après Schutzenberger, ne sera pas indispensable pour la 2ème Partie, où seules les informations de Fisher et de Shannon seront utilisées. Mais elle permettra d'introduire un concept de durée avec le processus de mesure, donc dès la base de la Physique : cette généralité initiale sera importante pour comprendre le rôle du temps en Thermodynamique et permettra de généraliser le démon de Maxwell.

0.3 - GÉNÉRALITÉS SUR LA DEUXIÈME PARTIE DU MÉMOIRE

0.3.1 - FONCTION DE GAIN ET CRITÈRES DES DIVERS JEUX

L'objet de notre jeu fondamental à trois joueurs de la Théorie des Communications sera de faire acquérir au récepteur la plus grande partie possible de la perte en "information" de l'émetteur, de façon à ce que la plus petite partie possible soit soustraite par la Nature.

Le récepteur étant le seul gagnant, c'est de son point de vue que la théorie doit être faite. Or ses actions sont nécessairement discrètes : elles consistent à agir ou ne pas agir, sans que l'on puisse rester entre les deux (Mais on peut admettre un certain pourcentage d'erreur dans l'action : de 1ère espèce si on agit quand on ne le devrait pas, de 2ème espèce si l'on n'agit pas quand on le devrait). Il se trouve que dans le cas discret, un seul concept d'information peut être considéré : c'est l'information sélective de Shannon et Wiener : c'est donc d'elle que seront dérivées les fonctions de gain,de risque et de coût de notre jeu.

On voit par ailleurs que la meilleure méthode de transmission de messages guidant des actions discrètes serait d'employer des signaux eux-mêmes discrets. Mais en fait la structure de la

Nature rend cela impossible, et de là vient un conflit entre aspects discret et continu, qui fait la richesse et la difficulté de la Théorie de la Communication et en particulier de la Thermodynamique.

Précisons que la "Nature", qui est simplement la source du bruit, ne doit être considérée comme un joueur que si l'on veut que le jeu soit à gain total nul : l'information apparente est perdue par l'émetteur, l'équivocation va à la Nature, et le reste au récepteur. On ne peut donc pas dire à priori que la Nature s'y prenne exprès de façon à démolir le message initial et diminuer la "valeur" du jeu (maximum de gain du joueur qui nous intéresse : le récepteur). Mais en fait, il se trouve à posteriori que c'est précisément ce qu'elle fait (§ 4. 3. 5.). Elle ne diffère d'un "vrai" joueur que parce qu'elle ne peut pas s'allier aux autres joueurs (mais son jeu peut tout de même dépendre du jeu des vrais joueurs, comme dans le bruit shot du § 4. 3. 5.).

(Si on accepte d'autres formes de Nature, on peut y inclure toute source de brouillage; ou l'ensemble Nature - Récepteur - Nature du double jeu du Radar, qui devient ainsi identique à un jeu simple).

L'information sélective étant donnée, comment déterminer d'abord les domaines dans lesquels varient les stratégies, puis le critère du jeu, enfin les coalitions qu'il comporte ? On ne peut pas les déterminer à priori, car on n'en connaît que les effets. Il faut donc procéder inductivement à posteriori. Le jeu à trois joueurs étant trop compliqué pour cette méthode, on considère trois jeux auxiliaires à deux joueurs :

<u>Emetteur - Nature</u>; <u>Récepteur - Nature</u>; <u>Emetteur - Récepteur</u>

On choisit des critères, qui ne sont à priori qu'assez vraisemblables. Ce sont respectivement :

- <u>le critère minimax</u> - <u>le critère Bayes</u> - <u>le critère minimin</u>.

appliqués respectivement à une <u>fonction de gain moins une fonction de risque</u> d'erreur (équivocation) pour le premier jeu et une <u>fonction de gain moins une fonction de coût</u> pour les deux derniers jeux.

Nous constaterons à <u>posteriori</u> que les théories correspondantes rendent compte numériquement de certains phénomènes observés : respectivement tout l'édifice de la Thermodynamique, la théorie du décodage, et les propriétés statistiques réelles des langues.

Nous appliquerons alors la méthode inductive fondamentale de la Phusique : de déclarer "<u>vrai</u>" un schéma dont les conséquences sont vérifiées (nos schémas seront très simples). Même si on n'accepte pas immédiatement cette conclusion, la concordance des résultats ajoute très considérablement à la vraisemblance de ces critères pour les phénomènes étudiés et encourage à continuer l'étude des modèles stratégiques en Physique.(Cependant, elle ne démontre en rien la validité de ces critères pour la description d'autres phénomènes).

Chacun des paragraphes § 0. 3. 2 à 0. 3. 4, puis des chapitres 4 à 6 sera consacré à l'un des jeux auxiliaires à deux

joueurs dont l'ensemble reconstituera un jeu à trois joueurs particulier (§ 0.3.4 discutera aussi cette reconstitution).

La fonction de gain ne dépend que de deux variables, on pourra à chaque fois appliquer la théorie de J. von Neumann et celle de nos chapitres 1 à 3; de plus, on se fera facilement une représentation géométrique des mouvements sur la surface de gain correspondant aux diverses opérations successives d'adaptation.

C'est des différences entre les critères qui permettent de retrouver les résultats expérimentaux des trois chapitres, que viendront surtout les différences entre les trois jeux, beaucoup plus que de la différence apparente entre leurs objets physiques (cependant celle-ci demeure assez importante du point de vue de l'aspect des problèmes et de leur public pour que les rédactions des divers chapitres soient gardées séparées).

0.3.2 - THERMODYNAMIQUE D'UN SIGNAL PARFAIT, CONFLIT ÉMETTEUR-NATURE, CRITÈRE DE MINIMAX

Le récepteur est supposé optimal et seule l'information sélective du signal émis est considérée, sa statistique étant laissée de côté. Ce signal est continu, mais sa connaissance doit servir à une action inductive, c'est-à-dire, doit transmettre un message discret. C'est là une nouvelle forme de l'opposition discret-continu que P. et T. Ehrenfest (1915) W. Pauli (1928) et R.C. Tolman (1938) ont traduite par l'opposition des entropies "petit grain" et "gros grain".

Les deux aspects discret et continu introduisent deux informations distinctes; dès la base de la théorie, et non comme superstructure ajoutée à posteriori. Ces deux informations sont également importantes, bien que la première, n'étant pas relative au récepteur, reste à l'arrière plan. Ce sont :

- l'information fishérienne, qui caractérise toute mesure du signal continu, car la stratégie de la nature a pour effet de limiter l'information fishérienne que peut acquérir l'expérimentateur à chaque mesure; cette limitation introduit le concept de température,

- l'information sélective, qui est la seule appropriée aux signaux discrets.

On supposera que la Nature veut minimiser et l'Emetteur maximer le gain en information sélective du récepteur, diminué d'une fonction de risque dans le décodage : l'équivocation. Il reste alors comme seules inconnues, les domaines de variation des stratégies. Du point de vue des joueurs actifs, l'étendue de ces domaines peut être traduite par une seule variable : l'entropie. Dès lors, on peut construire une "Thermodynamique d'un Signal parfait", redonnant les concepts et les lois de la Thermodynamique classique sans que l'on ait à attribuer au support du signal des propriétés précises telles que : gaz parfait, etc. D'ailleurs les résultats sont plus intuitifs dans le cas des ondes électromagnétiques. On a là un exemple de l'envahissement réciproque de l'électricité et de la thermodynamique.

Le critère minimax permet de considérer la Thermodynamique comme théorie de l'une des contraintes sur la nature des proces-

sus d'expérimentation, conséquences de la structure du monde physique. Il ne s'agira pas d'un nouveau "modèle" du type des Théories Cinétiques ou Statistiques, ni d'une "réinterprétation" informationnelle à posteriori. Nous étudierons aussi la théorie de quelques points voisins du minimax.

0.3.3 - THERMODYNAMIQUE DU DÉCODEUR. CONFLIT RÉCEPTEUR (ÉMETTEUR-NATURE) - CRITÈRE BAYES

Le message étant déjà envoyé, émetteur et Nature ne font qu'un joueur. On introduit le coût du décodage et on étudie sa relation avec la statistique du message, c'est-à-dire les probabilités relatives de diverses stratégies d'encodage de l'émetteur, dont la séquence constitue le message.

Pour le faire, on attribue aux mots des probabilités à priori et on minimise le coût probable à gain donné (ou maximise une fonction de gain-fonction de coût). On a donc affaire à une solution Bayes. Réciproquement la donnée du décodage détermine les probabilités des mots.

0.3.4 - THÉORIE DE LA LANGUE COALITION ÉMETTEUR-RÉCEPTEUR-CRITÈRE MINIMUM OU MINIMIN

On suppose que le décodage est séquentiel et optimum, ce qui permet d'éliminer la Nature en n'en conservant qu'un concept de coût. On étudie alors de façon formelle et détaillée la coalition de l'émetteur et du récepteur, régie par le critère de minimin de coût (minimum d'abord par rapport au récepteur, puis à l'émetteur).

Cette coalition conduit à une distribution canonique que devraient suivre les éléments du message : si on les range par ordre de fréquences décroissantes, il faut que la probabilité de l'élément numéro n soit :

$$p_n = P\,(n + m)^{-B} \qquad \text{(cf. fig. 1, p. \ )}$$

(Le paramètre m est la seule fonction des détails du codage : il résulte du fait que celui-ci est arithmétique).

La statistique canonique tire son importance du fait qu'il semble bien que la structure statistique de toutes les langues soit canonique, lorsque l'on prend comme éléments les mots, pris sous forme entièrement fléchie (et non comme les unités du lexique).

C'est là le fait nouveau le plus important de ce mémoire. Il résulte par exemple des données réunies par Zipf (1949) dont certaines sont reproduites fig. 3 pp.

Signalons que Zipf pensait pouvoir se contenter d'une formule sans paramètre : $f_n = P/n$, à laquelle il apportait ensuite des corrections empiriques, pour tenir compte des nombreuses divergences. (Nous ignorions ces divergences en construisant notre théorie, car nous ne connaissions alors qu'un résumé du livre de Zipf, qui ne les signalait pas). Au contraire, la loi canonique justifie automatiquement, et de façon rationnelle, et non pas par coups de pouce empiriques, aussi bien la pente différente de

– 1 pour n grand (en coordonnées bilogarithmiques) que la divergence par défaut pour n petit.

Ainsi, moyennant l'introduction de coalitions appropriées, une méthode faite pour étudier des problèmes de conflit permet de résoudre un problème d'organisation, apportant ainsi un des premiers résultats de la Physique de la Langue, science nouvelle en train de naître (Jakobson 1951). (La signification linguistique de ce résultat est étudiée en détail dans une autre publication). C'est la méthode stratégique, ainsi que le caractère "immotivé", "arbitraire", du signe linguistique, sur lequel insiste Saussure, qui justifient l'omission intentionnelle dans cette théorie de toute référence aux propriétés de l'émetteur et le fait que la théorie des messages émis est tout entière axée sur le récepteur.

La concordance entre loi théorique et données expérimentales cesse pour les textes très artificiels en "Basic English" ou en "Espéranto".

Elle est au contraire particulièrement frappante dans le cas de certains schizophrènes (ceci est très naturel étant donné la très faible "cohérence sémantique" des textes qui leur sont dus, dont le sujet change constamment : l'exemple extrême de schoziphrénie serait donné par les "textes" pour enfants où chaque groupe de syllabes est en même temps la fin d'un mot et le début du suivant).

<u>Notre conclusion est que,lorsqu'il peut être défini, le mot constitue un morceau naturel d'information adapté au codage arithmétique</u>.

La théorie rendra par suite possible, et même très simple, la description mathématique de certains aspects de la langue, comme nous le verrons au Chapitre 7. Mais elle ne donnera pas un sens opérationnel à tous les paramètres que les linguistes ont considérés. En particulier, elle montrera que dans la majorité des cas, la notion de "nombre de mots potentiel" est dépourvue de sens. Par contre, pour les paramètres pourvus de sens, elle donnera des méthodes rigoureuses de calcul, à partir des données empiriques.

Nous proposons d'ajouter la propriété d'être canonique aux "propriétés" par lesquelles le linguiste Ferdinand de Saussure "définissait" les "entités concrètes" de la langue. Cette définition transforme en loi expérimentale la présomption de Saussure, que le <u>mot est l'entité concrète de la langue</u>.

(Il n'y a aucune raison à priori de conclure que le mot est la seule entité concrète dans notre sens. Il peut y en avoir beaucoup d'autres, plus courtes ou plus longues, sans qu'elles puissent d'ailleurs être trop courtes car elles seraient alors trop peu nombreuses pour qu'il soit possible d'en donner une description canonique ne faisant pas intervenir les détails précis du codage : Cf. § 6.2.1).

Cependant, dans certaines langues, l'unité naturelle d'information n'est pas aussi nettement définie que dans celles étudiées par Zipf. Alors, par application d'un processus de va-et-vient nous pourrons dire que la "meilleure", à l'intérieur d'un groupe de définitions possibles du mot,sera celle que donnera la

loi la plus proche de notre loi fondamentale. (Toutefois, étant donné le caractère "statistique" du critère de choix, il ne peut pas permettre de distinguer entre définitions trop voisines).

Il existera très certainement encore de nombreuses exceptions à la règle d'optimum. Il nous faut donc souhaiter que les décomptes se multiplient. Espérons qu'il sera possible, soit encore d'expliquer les exceptions, soit de les cantonner à des textes caractérisables par des traits macroscopiques "pathologiques". Si ceci se révèle impossible, la théorie devra bien entendu être profondément révisée.

Que peut-on induire du caractère canonique des mots ?

Il semblerait inconséquent de refuser d'appliquer dans ce domaine la méthode inductive fondamentale de la Physique : Nous attribuerons donc le caractère canonique des mots au fait qu'il est "vrai" que la langue est construite de façon à ce que l'information sélective puisse être décodée à moindres frais par un décodeur mot-par-mot optimum.

Pourquoi chercherait-on cependant à rendre optimale l'organisation du niveau morphologique, alors que le centre du problème de codage de la langue se situe plus près du codage sémantique que du codage morphologique ? Pour y répondre, il faut remarquer que la stratégie de transmission de l'information est hiérarchisée. Le but de cette hiérarchisation peut être cherché dans le désir de pouvoir limiter l'activité consciente au niveau sémantique, et être sûr que l'optimum est atteint au niveau morphologique, où les servitudes physiques sont les plus grandes. Pour pouvoir négliger ces servitudes et se livrer librement, au niveau sémantique, aux multiples coalitions dépendant du but et des caractères de leur conversation, les interlocuteurs conviennent donc de s'allier au niveau morphologique. Cette convention se fait progressivement par le choix de la langue que l'on parle, tout au long de son apprentissage : ensuite les deux joueurs jouent de la façon la plus agréable pour celui qui reçoit le message (cf. 0.1.1.).

C'est la troisième fois que nous avons précisé le critère d'un jeu à deux joueurs,mais pour que ce critère soit satisfait, il n'est pas nécessaire que les deux premiers critères le soient aussi. En effet, du fait que nous avons dès le début de ce § 0. 3.4 éliminé la Nature, il résulte qu'un jeu émetteur-récepteur optimal est compatible avec des jeux émetteur-nature et nature-récepteur non optimaux. En d'autres termes, l'organisation des opérations peut être optimale, sans qu'elles le soient elles-mêmes. D'ailleurs, tel est évidemment le cas dans l'exemple linguistique, comme dans l'exemple du délai dans le processus de reconnaissance, étudié par R. W. Hick (6.2.2). Dans ces deux exemples,comme dans tous les autres problèmes réels de transmission, si les opérations élémentaires étaient elles-mêmes optimales, les chances d'erreur seraient excessives.

Le fait que les phénomènes concordants se produisent pour le langage et le temps de réaction n'a rien de surprenant, étant donné qu'il s'agit là de deux processus psychologiques, de très "haute" localisation cérébrale. Par ailleurs ce fait augmente notre confiance dans l'explication de la langue par un critère de minimin.

PREMIÈRE PARTIE

CHAPITRE I

COMPORTEMENT INDUCTIF

1.1 - LE CONCEPT DE COMPORTEMENT INDUCTIF DE J. NEYMAN

1.1.1 - DÉFINITION

J. Neyman a introduit le concept du comportement inductif vers 1930, et l'a décrit (1947) dans les termes suivants :

" On a exprimé l'opinion que le calcul des probabilités et " la statistique mathématique forment la base d'un procédé men- " tal qu'on nomme "raisonnement inductif". Si l'on convient à " n'employer le mot "raisonnement" qu'à la description des pro- " cédés qui conduisent au savoir, le raisonnement ne saurait " être que déductif. D'autre part, comme il a été signalé (par " Neyman 1937), il y a lieu d'employer le terme "comportement " inductif". On entend par ce terme les règles, plus ou moins " générales, d'ajustement de nos actions aux résultats d'obser- " vations limitées.

" Le calcul des probabilités et la statistique mathématique " jouent un rôle important dans la formulation de ces règles. " Certainement il y a assez de raisonnements, mais, comme d'ha- " bitude, tout raisonnement est déductif ".

La correspondance entre action et résultat des expériences établie par la règle de comportement inductif s'exprime par la "fonction de décision statistique" de la règle. Sa recherche constitue le problème de la construction du plan d'expérimentation correspondant à l'ensemble des hypothèses admissibles. Ce problème a été posé de façon générale par A. Wald (1950). Pour lui, le statisticien est un être calculateur en conflit avec la Nature. Celle-ci est représentée par un processus stochastique, c'est-à-dire une Urne généralisée, ce qui peut être considéré comme résultant de l'inversion d'une expression de Quételet : "L'Urne que nous interrogeons, c'est la Nature".

Cette Nature est inerte, c'est-à-dire qu'on ne peut pas dire qu'elle cherche à nuire au statisticien. Le problème est donc, strictement parlant, intermédiaire entre le "conflit" entre deux êtres non calculateurs et le duel stratégique entre êtres calculateurs. Si cependant l'on désire ne jamais pêcher par excès d'optimisme, il faut que la règle de comportement soit

valable en toute circonstance, ce qui exige de supposer que la Nature joue effectivement contre le statisticien. Donc le problème statistique de Wald se ramène à un "duel" pur : en fait, la technique mathématique de la Stratégique s'est finalement révélée plus fructueuse dans ce problème que dans le problème économique initial, où elle reste quelque peu qualitative.

(Le raisonnement précédent qui sert à justifier le critère de "minimax" est souvent attaqué par des raisons à priori. Nous allons cependant montrer au chapitre 4 qu'il conduit à posteriori aux résultats corrects et expérimentalement vérifiables dans le cas de la Thermodynamique, tandis qu'aucun autre critère simple ne semble y conduire).

1.1.2 - PASSAGE DE LA STATISTIQUE A LA THÉORIE GÉNÉRALE DE L'EXPÉRIENCE EN PHYSIQUE

La théorie des fonctions de décision a surtout été étudiée en contexte purement statistique. Mais il est évident que rien d'essentiel ne distingue du Statisticien l'Expérimentateur le plus général.

Par exemple, à la fixité de recettes statistiques correspond la fixité des instruments de mesure, et les unes et les autres peuvent être utilisées sans que l'on comprenne la raison de leur fonctionnement.

La seule différence, mais notable, est que l'action du statisticien est souvent, ou bien négligeable, ou bien destructrice tandis que des lois physiques déterminent l'action minimum de l'expérimentation ; ces lois peuvent d'ailleurs inversement être interprétées par le "dérangement de l'objet par la mesure".

Cette analogie Statistique-Expérience est bien familière. Elle se juxtapose à celle Stratégique-Statistique. Il est tout indiqué de voir si l'utilisation directe de la Stratégique, sans passer par l'intermédiaire statistique (comme le fait Fisher) ne pourra pas rendre le même service à l'expérimentateur qu'au statisticien : lui donner une pleine conscience des conséquences des mouvements (opérations physiques effectives) dont l'ensemble constitue le mode opératoire fourni par la Physique pour vérifier si une relation est vérifiée ou non.

Pour nous, la suite de toutes les opérations effectuées par un expérimentateur pour déterminer l'état d'un système se réduira à un seul mouvement : choix des instruments et du mode opératoire. Ils resteront fixes une fois choisis et constituent la stratégie d'expérimentation. La technique de von Neumann (1944), réduisant tout jeu au choix initial de la stratégie, se trouvera donc aussi parfaitement adaptée au problème, que l'est le concept fondamental de stratégie lui-même. (On peut également considérer la stratégie comme une suite de transformations de l'objet, se combinant suivant des lois à déterminer).

Le problème de l'expérience est très bien posé dans l'important ouvrage de Lawson et Uhlenbeck : Threshold Signals (1950). Ceux-ci "font remarquer (p. 168, note) que la théorie de l'observateur idéal est pratiquement identique à la théorie (due à Neyman et Pearson) du meilleur critère pour le test d'une hy-

pothèse statistique". Ailleurs, ils font remarquer que trouver un signal dans du bruit est un problème de jeu où intervient le hasard.

Le travail de Lawson et Uhlenbeck a été continué dans l'important mémoire de D. Middleton (1952), qui est parvenu à notre connaissance lorsque ce travail était déjà rédigé. Nos deux tentatives se complètent d'ailleurs, Middleton ayant traité des problèmes directs, et nous-mêmes surtout des problèmes inverses, au sens de la classification qui va être exposée dans le paragraphe suivant.

1.2 - CLASSIFICATION DES PROBLÈMES DE COMPORTEMENT INDUCTIF

1.2.1 - DÉFINITIONS

En reprenant la définition de Neyman, nous dirons tout-à-fait généralement que l'objet de la théorie des fonctions de décision est de réaliser l'adaptation entre la stratégie et son objet propre, en tenant compte de toutes les contraintes qu'impose la nature du monde physique.

Par "objet" l'on entend, bien entendu, un élément inconnu d'un processus stochastique supposé connu. La stratégie sera donc adaptée au processus. Nous verrons au § 1.2.5 que l'adaptation dépend également du nombre de fois que la stratégie doit être appliquée au processus (ce nombre n'a rien de commun avec le nombre d'opérations d'une stratégie séquentielle, que l'on doit d'abord laisser indéterminé : il est l'analogue du nombre de fabrications à étudier par la méthode séquentielle). Mais nous négligerons d'abord ce nombre en le supposant fixe, et n'aurons donc affaire qu'à trois éléments :

- CRITÈRE D'ADAPTATION, y compris les domaines de stratégie
- OBJET,
- STRATÉGIE.

Si un seul de ces trois éléments est inconnu, on doit pouvoir le déduire des deux autres, tout au moins conceptuellement. Introduisons la terminologie suivante :

- Si l'objet et le critère sont donnés, le problème sera dit DIRECT;
- Si la stratégie et le critère sont donnés, le problème sera dit INVERSE;
- Si l'objet et la stratégie sont donnés, le problème sera dit MÉDIAN.

Nous allons exposer d'abord pourquoi il faut introduire les deux dernières catégories de problèmes, qu'on ne rencontre guère "directement" dans les problèmes que pose la pratique.

1.2.2

Un problème direct se présente en général sous la forme de l'organisation des opérations connues et permissibles en vue d'un certain objet. En fait, très peu nombreux sont les cas où la stratégie optimum peut être obtenue, ne serait-ce qu'à cause

des difficultés mathématiques, ou à cause des difficultés qu'il y a à décrire un objet de façon suffisante pour la stratégie, et en même temps non surabondante.

Lorsque les difficultés sont de cette deuxième espèce, on peut les contourner de la façon suivante. : 1° : on construit un ensemble de stratégies particulièrement simples, et dépendant de peu de paramètres; 2° : on résoud les problèmes inverses correspondants; 3° : on identifie l'objet que l'on possède à l'un des objets-étalons numérotés.

La méthode inverse s'imposera donc à chaque fois que l'on sait que le nombre de stratégies possibles est limité et par suite la stratégie susceptible d'être spécifiée par un petit nombre de propriétés.

En somme, pour utiliser une comparaison, pittoresque mais frappante, le progrès que constitue l'utilisation de cette méthode est identique à celui que constitue le passage du tailleur sur mesures au tailleur de confection : chacun sait qu'il est plus facile d'indiquer une taille de confection que d'énumérer les dimensions d'un individu, et que l'habit est ensuite plus vite disponible.

Les résultats des deux méthodes seront cependant identiques dès que la collection sera assez riche et le client assez "normal". Tel sera fréquemment le cas en Physique. Mais même dans le cas opposé, la méthode inverse peut permettre de faire une partie du choix, en ramenant au choix initial des méthodes pour ce qui concerne les détails.

1.2.3.

En fait, nous allons restreindre notre définition des problèmes inverses, en introduisant la condition que les stratégies-étalon choisies soient elles-mêmes les meilleures, dans un sens déterminé, c'est-à-dire résultent de la solution de certains problèmes directs particulièrement simples (nous verrons en détail au § 6.3.3 que le produit des opérations directe-inverse ne ramène pas toujours à l'objet initial).

On peut aussi considérer comme inverses les problèmes que pose l'étude des propriétés des stratégies "les meilleures" dans un sens approprié au but de ces stratégies, et des objets adaptés à ces stratégies; et rattacher aux problèmes directs l'étude d'une stratégie ou d'un objet non optimaux ou définis indépendamment de tout critère, par exemple par un mécanisme réel. Mais même si l'on procède de cette façon, l'introduction d'un optimum dans la définition des problèmes inverses fait que notre classification ne constitue plus une vraie dichotomie.

Toutefois, elle se révèlera très utile en tant que principe classificateur applicable de nombreuses fois successivement dans un même mode d'étude. Parmi les distinctions qu'elle rendra inutiles, citons :

- la distinction entre analyse et synthèse dans certains contextes. Par exemple l'analyse de l'onde acoustique de la parole peut être assimilée à un problème direct. La synthèse de la parole à l'aide d'éléments connus participe de l'"esprit inverse"

- la distinction mathématique entre problèmes constructifs et d'existence (axiomatique) tout au moins dans ce contexte physique. En fait, la dichotomie physique garde de la dichotomie mathématique la fâcheuse propriété que bien fréquemment la stratégie, dont un problème inverse montre l'existence, n'est pas réalisable, parce qu'elle comporte une limite, sortant du domaine de réalisabilité considéré.

- la distinction inverse-direct se retrouve aussi dans les questions terminologiques : exemple : "Sémiologie" est un terme inspiré de préoccupation directe, et "Théorie de l'Information" est inspiré de préoccupation inverse.

Par ailleurs, la méthode directe est typiquement une méthode de découverte; la méthode inverse, une méthode d'exposition.

1.2.4 - PROBLÈMES MÉDIANS

On n'a affaire ni à un problème direct, ni à un problème inverse, dans les cas où l'on suppose, pour des "raisons" somme toute irrationnelles, qu'une adaptation est réalisée, sans que l'on sache exactement laquelle. Un cas typique est celui où l'on divise un système physique inerte en évolution en deux parties, et appelle stratégie l'action de l'une sur l'autre. Le centre de gravité de la théorie s'est alors déplacé vers l'étude du critère de cette adaptation. L'égalité de l'action et de la réaction serait un tel critère, mais nous montrerons qu'il existe aussi des critères plus complexes, par exemple du type minimax = hostilité complète de l'adversaire, ou minimin = coopération.

Il ne semble pas qu'il faille introduire ces nouveaux problèmes pour élargie la "dichotomie" direct-inverse, car techniquement leur forme est presque déserpérée. Il faudra donc dans chaque cas chercher un moyen de les aborder différemment.

Par exemple, on peut choisir le biais de l'adoption provisoire d'un critère, suivie du choix de l'une ou l'autre des méthodes fondamentales de la dichotomie. On peut en principe choisir l'une ou l'autre indifféremment. Cependant, en pratique, il se trouvera que la variété fonctionnelle des stratégies est plus petite que celle des objets et on sera tenté d'adopter plutôt la méthode inverse.

Il en résulte que nous n'aurons à nous occuper par la suite que de problèmes physiques inverses. Mais l'étude de ceux-ci n'a d'intérêt et d'importance pratique que par rapport aux problèmes directs qu'elle contribue à résoudre et dont elle tire le plus souvent le plus clair de son inspiration (même le contexte philosophique de ce travail est dû à l'étude de nombreux problèmes directs). Nous les garderons donc constamment présents à l'esprit et aurons fréquemment à en rappeler certaines solutions.(La suite de ce § 1.2 et le § 1.3 constituent une parenthèse méthodologique, que l'on peut sauter sans grave inconvénient pour la suite des idées).

1.2.5 - ERGODICITÉ

Les hypothèses et théorèmes ergodiques de la mécanique statistique sont relatifs à des égalités entre une moyenne à temps constant et lieu variable, et une moyenne à lieu constant et

temps variable. Prenons maintenant le terme "ergodique" sous la forme la plus générale comprise dans cet énoncé. Alors, on voit que toute relation d'"ajustement" ou d'"adaptation" des actions aux expériences est une relation ergodique, entre une "moyenne" prise sur la stratégie à n'importe quel instant, c'est-à-dire sur un ensemble direct réel, et une "moyenne" prise sur l'ensemble des réactions admissibles, c'est-à-dire sur un ensemble inverse conceptuel.

Mais l'égalité ainsi comprise ne peut pas être un théorème mais seulement une condition qu'il doit être nécessaire de réaliser. En fait, cette condition sera souvent réalisée par une majorité de stratégies d'un certain ensemble, donc on aura comme théorème l'égalité entre deux moyennes toutes deux conceptuelles. On peut aussi avoir un théorème en prenant une limite de moyennes réelles qui se révèle égale à une moyenne conceptuelle. Mais les deux termes sont alors tous deux conceptuels. (Exemple: théorie de Gibbs).

1.2.6 - STRATÉGIES HIÉRARCHISÉES ET INFLUENCE DU NOMBRE TOTAL D'OPÉRATIONS A EFFECTUER

Replaçons nous dans le cadre du problème direct général de la construction de la stratégie. Peut-on déterminer d'avance si elle sera hiérarchisée, c'est-à-dire constituant une organisation de sous-stratégies (elles-mêmes éventuellement décomposables), ou si elle ne sera pas hiérarchisée, mais "une en soi" ?

On constate que dans la réalité, toutes les stratégies (et en particulier les stratégies "naturelles" des êtres vivants) sont très fortement hiérarchisées : Ce fait résulte-t-il de ce que ces organismes sont incapables de considérer simultanément un groupe trop important d'opérations, et doivent donc hiérarchiser pour se ramener à une organisation optimale d'un nombre plus petit de groupes d'opérations ? Ou bien au contraire peut-on atteindre exactement l'adaptation avec des stratégies hiérarchisées ?

La réponse à cette alternative introduit le nombre total d'opérations à effectuer, que l'on avait laissé de côté au début du § 1.1.1. En effet, le problème se ramène au suivant : à la stratégie hiérarchisée que l'on considère correspond par adaptation une certaine population. Le nombre d'épreuves dont on dispose permet-il ou non de la distinguer de la population donnée à étudier?

Prenons un exemple : l'on effectue des tirages dans une urne dont les boules se rangent en classes de probabilités égales. On classe les boules sorties par fréquence décroissante. S'il n'y avait pas dispersion, la courbe obtenue serait une courbe en escalier, chaque marche correspondant à une classe. Mais la dispersion amortit en réalité ces marches, tant que le nombre de tirages ne dépasse pas une certaine valeur, fonction de l'écart maximum entre deux classes voisines.

Donc une population à classes (qui se contente de stratégies hiérarchisées : peu d'opérations à grand rendement) ne peut pas être distinguée d'une population à classes réduites à une

boule chacune (qui aurait exigé une stratégie adaptée à beaucoup d'opérations de faible rendement chacune).

Par suite, l'on peut toujours adopter une stratégie hiérarchisée, mais qui doit l'être d'autant plus finement que le nombre d'épreuves est plus grand. Ce n'est que pour un nombre d'épreuves infini qu'il faudra une stratégie non hiérarchisée, "une en soi".

Ceci est d'ailleurs conforme à l'intuition, mais précise celle-ci en montrant que pour un nombre d'utilisations donné, on n'améliore plus la stratégie en augmentant sa complication au delà d'une certaine limite, qui croît avec le nombre d'utilisations.

Ce résultat a par exemple grande importance dans la théorie des machines à calculer, où les sous-routines doivent être d'autant plus nombreuses et fines que le nombre d'opérations à effectuer est plus grand.

Ce résultat permet aussi d'enrichir le concept d'"ergodicité" du § précédent de la considération réaliste des "moyennes temporelles" finies. Chaque cas où il n'y a pas ergodicité avec moyennes infinies définit une longueur de temps au-dessous de laquelle cette absence d'ergodicité ne peut être constatée. Cette longueur de temps donne ainsi l'ordre de grandeur du temps au bout duquel il peut devenir souhaitable de réadapter la stratégie. (cf. aussi § 1.3.6 fin).

(Ce résultat est enfin intéressant pour la considération de machines qui s'organiseraient elles-mêmes de mieux en mieux d'après les défauts de fonctionnement qu'elles constateraient. Elles s'adapteraient de plus en plus finement, et deviendraient "unes" au bout d'un temps infini : il n'y aurait plus alors d'"organes" individualisables).

1. 3 - PROBLÈMES DE COMPORTEMENT DE LA PHYSIQUE

1.3.1 - PROBLÈMES PHYSIQUES

L'un des principaux avantages de la distinction entre :

- problèmes physiques directs,
- problèmes physiques inverses,
- problèmes physiques médians,

est qu'elle permet de préciser, et à vrai dire d'éliminer, la vieille distinction entre :

- problèmes techniques,
- problèmes physiques,
- problèmes physiques explicatifs (modèles).

La distinction entre les deux premiers n'est qu'une forme de la distinction entre sciences pures et appliquées, qui est tout-à-fait artificielle du point de vue réel de ceux qui pratiquent la science.

En fait, les plus typiques des problèmes directs sont ceux que posent les décisions inductives de la vie quotidienne et de la technique qui en est le prolongement. Employer la méthode di-

reste de bout en bout est en général un trait d'imprévoyance, car elle comporte des difficultés particulières.

Les plus typiques des problèmes inverses sont ceux que pose l'étude des limites que la structure du monde impose aux actions, donc les problèmes de la "science anthropocentrique" : celle qui ne se propose pas d'expliquer les limites aux stratégies, mais les établit en vue de problèmes directs. L'existence de cet aspect de la Physique contribue à faire comprendre pourquoi la Physique est utile à la technique.

(En pratique, tout exposé cohérent de ces limites ne s'appuiera pas sur des expériences réelles, qui auraient exigé de résoudre un problème direct de façon optimum, avant d'en connaître les limites; mais il se basera sur des "expériences conceptuelles".)

Le recueil des solutions des problèmes inverses constitue l'élément de prévoyance dans la solution des problèmes quotidiens ou techniques. C'est parce que la condition d'optimum s'introduit naturellement dans les problèmes inverses de la Physique, qu'elle a été introduite dès le début du § 1.2.

1.3.2 - PAS DE PHYSIQUE DIRECTE OU INVERSE

Bien que la distinction entre problèmes physiques directs et inverses soit à notre avis fondamentale du point de vue de la méthode physique, il n'y a rien de tel qu'une "physique directe" ou une "physique inverse".

Tout d'abord, un même problème peut être analysé en éléments dont les uns sont directs, les autres inverses. Des problèmes d'apparence directe font intervenir les solutions des problèmes inverses, à la fois comme limites aux combinaisons nouvelles possibles, en cours d'étude, et comme simplifications de solutions de problèmes directs presque parfaits, étudiés antérieurement.

D'autre part, un problème direct ou inverse par rapport à un niveau d'étude peut avoir des aspects multiples par rapport à un autre niveau, et ceux-ci peuvent être inverses l'un par rapport à l'autre. Par exemple :

1°) La synthèse et l'analyse de circuits sont les aspects direct et inverse d'un problème direct, dont le problème inverse est résolu par les lois physiques de l'électricité.

2°) La synthèse et l'analyse d'une langue réelle ou artificielle sont les aspects direct et inverse d'un problème inverse, celui qui mène à la définition de l'information.

Il ne faut jamais confondre les concepts relatifs aux problèmes directs et inverses (cf. par exemple § 5.4), mais les solutions de certains problèmes inverses peuvent s'obtenir à partir de solutions de problèmes directs en faisant disparaître progressivement des paramètres introduits par ces derniers (cf. Gabor - 1950). (Si la disparition a lieu par passage à la limite, la solution directe peut cesser d'en être une). Nous utiliseront un procédé de cette espèce pour définir certaines classes de stratégies.

Certains résultats importants de la science exprimeront le fait qu'une notion d'origine inverse donne le même résultat que la limite d'une notion directe. Ce sera le cas pour le théorème du codage sans bruit de Shannon : l'information sélective s'obtient à la fois à posteriori (inversement) comme solution d'un système d'axiomes, et à priori (directement) comme limite du coût de transmission dans une classe de systèmes, dont les limites correspondent aux conditions additionnelles qui rendent déterminée la solution des axiomes.

1.3.3 - PHYSIQUES A PRÉDOMINANCE INVERSE

S'il n'y a pas de "physique directe" et de "physique inverse", il y a cependant des branches de la physique dont les lois sont à prédominance directe ou inverse. Dans l'intérêt de la cohérence logique de la science, il y a alors intérêt à exagérer ce caractère dans leur exposé.

Par exemple, la thermodynamique macroscopique phénoménologique classique est un recueil de solutions de problèmes inverses. Nous nous efforcerons d'accentuer encore ce caractère dans l'exposé que nous en ferons au Chapitre 4, exposé où les concepts d'information apparaîtront dès la base. Par suite de cette circonstance, une expression de l'entropie formellement identique à celle de Boltzmann y jouera un rôle important : mais le point de vue appartiendra cependant intégralement à la "théorie inverse" et non pas aux modèles macroscopiques, concrets (Théorie cinétique) ou abstraits (mécanique statistique). Il ne se prétendra pas explicatif, mais réinterprétera les limitations phénomènologiques comme aspect de la durée.

C'est uniquement à ce caractère "inverse" que tient l'aspect très particulier que présente la Thermodynamique parmi les sciences. Cet aspect a été remarqué par de nombreux auteurs (cf. **Bridgman**, 1943); il n'est cependant pas propre à la thermodynamique, qui le partage avec tous les principes de symétrie ou d'antisymétrie, avec le principe d'exclusion de W. Pauli et avec les lois physiques qui servent de base expérimentale à la relativité : 'lois de Michelson, de Kennedy et Thorndike et de Ives et Stilwell. - Cf. H.P. Robertson, 1949).

Les plus utilisées des solutions de problèmes inverses sont celles qui se traduisent par la simple interdiction de toute acquisition d'information d'une certaine catégorie sur un certain ensemble. La simplicité de cette forme est due uniquement au caractère approprié du cadre conceptuel stratégique et du formalisme mathématique qui le traduit. On est en droit de considérer que la question de savoir si toute la physique peut être réduite à une telle forme est l'une des questions importantes que la Physique ne pourra jamais résoudre, mais que l'on doit tenir présentes à l'esprit.

1.3.4 - PERMANENCE DES RÉSULTATS

Il y a, entre les solutions de problèmes physiques directs et inverses une grande différence de "permanence".

Si la nature du monde ne change pas, les solutions de problèmes directs gardent toujours la même validité. Au contraire, les solutions de problèmes inverses sont en dernière analyse

d'origine expérimentale, et les progrès de la physique contribuent constamment à les modifier.

Cependant, l'utilité pratique des problèmes directs diminue, à validité constante, tandis qu'augmentent les exigences techniques du constructeur, tandis que l'utilité des problèmes inverses reste définitivement acquise.

La situation idéale serait celle où problèmes directs et inverses seraient confondus du point de vue du résultat et où tout problème technique serait tout résolu dans les manuels de physique. Mais comme ce n'est pas le cas, il faut considérer séparément les deux problèmes "accidentellement" confondus.

1.3.5 - PHYSIQUE EXPLICATIVE

En un certain sens, les problèmes médians englobent les problèmes physiques explicatifs (dont l'existence était ce qui rendait ambigü le terme "physique") : ce sont ceux qui concernent les modèles non anthropocentriques des contraintes physiques (et remettant ainsi l'homme à sa place). Ces modèles gardent d'ailleurs rarement jusqu'au bout leur caractère explicatif : par exemple, le développement du modèle quantique a rendu nécessaire le principe d'exclusion de Pauli, qui est une limite ayant typiquement la forme de la solution d'un problème inverse.

Le problème de la transmission est également médian : il s'agit de déterminer un critère d'adaptation et une information par rapport à un problème direct résolu : l'organe des sens. Les problèmes médians sont également l'aspect le plus apparent du problème que posent les techniques utilisées par des êtres vivants, si elles sont très stables et dont, par suite de leur long rodage avant d'arriver à cette stabilité, on peut croire qu'elles sont en un certain sens les meilleures possibles. (Ex : instinct, réflexes...). Alors le problème direct ne se pose pas, puisque la stratégie est connue, et l'inverse est inutile. Mais la stratégie et l'objet sont en général mal définis et difficiles à décrire exactement. On veut alors vérifier si ce que l'on sait est compatible avec le critère supposé. Si oui, ceci ajoutera, et très considérablement, à la confiance que l'on aura dans chacun de ces trois éléments séparément. Nous étudierons en détail un tel exemple dans le cas de la structure des langues.

1.3.6

Puisque les problèmes médians se ramènent à des problèmes directs ou inverses, il faut indiquer comment les deux attitudes se présentent dans le problème de la physique explicative, compris comme étant un problème de correspondance entre signes et phénomènes.

L'attitude directe part de ce qui est donné, le phénomène, pour le représenter au moyen de signes et de leurs règles de combinaison. Le codage par les résultats d'expérience constitue un "résumé exhaustif" (Fortet 1950, p. 237) du système physique étudié, dans le sens généralisé d'ensemble des valeurs des paramètres indépendants, sans intervention nécessaire du hasard (Un tel résumé est strictement conforme à la réalité physique si le signal est un système au sens de la mécanique rationnelle, dont les paramètres sont déterminables rigoureusement. Dès qu'inter-

viennent le hasard et les fluctuations, il faut considérer le résumé exhaustif au sens propre, résultat d'un grand nombre d'expériences systématiques ou non, indépendantes ou non). L'attitude inverse cherche à retrouver ce qui est donné à partir de signes.

Le problème direct est celui de la Physique théorique. L'attitude du savant est celle du décrypteur. Le problème inverse est celui de la Physique mathématique ou de la Mathématique appliquée. L'attitude du savant y est celle de l'encodeur.

Pour effectuer le décryptage, il faut bien entendu posséder les résultats d'expérience et aussi l'expérience de systèmes de code aussi nombreux que possible. Pour cela, il faut résoudre de nombreux problèmes inverses qui sont en principe moins difficiles que les directs. Il n'y a bien entendu aucun espoir de résoudre le problème direct général par une attaque frontale : même les problèmes directs spéciaux sont difficiles et ne sont solubles que si la correspondance idée-expérience est presque biunivoque. D'où la nécessité d'une étude zigzagante, faite au hasard ou de façon exhaustive.

Le physicien mathématicien cherche à multiplier le nombre de concepts et structures dont il tient compte dans son édifice. Le physicien théoricien au contraire cherche à le diminuer. Cependant, le deuxième fait toujours intervenir plus de concepts que le premier. Ce fait exprime l'imperfection de la Physique, et la dualité entre signes et phénomènes qui se reconstitue après chaque courte période où la Physique semble être achevée.

Quel peut être l'aboutissement de cette complication progressive ? Faisons à ce sujet une remarque, relative aux stratégies hiérarchisées. La Physique explicative peut être considérée comme une stratégie adaptée aux expériences passées, et que l'on présume par suite adaptée aux expériences futures. Mais le nombre des expériences passées agit ici de la même façon que le nombre des expériences futures le faisait au § 1.2.6. Si ce nombre était petit, la stratégie pouvait être hiérarchisée, sans rien perdre de son efficacité. De même ici, la Physique peut être constituée de disciplines disjointes, sans rien perdre de son efficacité.

Mais quand le nombre d'expériences croissait, la Stratégie pouvait tendre, sans d'ailleurs le faire nécessairement, vers une stratégie "une en soi", sans sous-stratégie autonome. De même, il se pourrait fort bien, sans que cela se produise nécessairement, que, le nombre d'expériences passées croissant indéfiniment, il devienne rigoureusement impossible d'attribuer le moindre sens à l'idée de "discipline physique", tout phénomène ne pouvant être expliqué de façon satisfaisante qu'en faisant intervenir toute la Physique.

1.3.7 - LES THÈSES FONDAMENTALES DE LA PHYSIQUE

Une classification quelconque n'a de sens et d'utilité que si elle n'est pas vide et en particulier la thèse fondamentale de la physique est qu'il existe des problèmes solubles de toutes les catégories ci-dessus.

Pour voir ce que signifie cette thèse, nous allons l'introduire progressivement à partir d'une antithèse fondée sur une vieille théorie physiologique.

Thèse d'Empédocle : On ne peut reconnaître de la lumière et du son, que si l'on possède dans l'organe de sens de la lumière et du son de même nature physique.

Bien entendu, ceci n'a maintenant aucune valeur : du point de vue moderne, la "nature physique" est quelque chose de pas clair du tout : elle ne paraissait claire à Empédocle que parce qu'il ne s'agissait pour lui que de l'un des quatre "élements", eau, feu, air, terre.

Cependant, nous allons reprendre, moderniser et généraliser l'idée centrale de cette thèse, car elle représente très exactement l'antithèse de la Physique telle que l'on la conçoit maintenant :

Antithèse fondamentale de la Physique : Il est impossible de caractériser la relation entre deux objets physiques autrement que comme différence ou identité.

Il peut paraître inutile à un physicien de dénoncer de nouveau ce principe macroscopiquement absurde. C'est cependant à cela précisément que se réfèrent les "médecins" qui disent : "il n'y a pas de maladies, il n'y a que des malades". L'idée qu'il n'existe que des objets physiques dont la seule relation peut être l'identité, est par suite encore très vivace. Pour elle, il n'y a pas d'identité partielle, il n'y a ni notions de grandeur physique, ni mesure, ni structure physique : on a une topologie discrète pour l'ensemble des objets physiques.

Le principe d'Empédocle caractérise le chaos : En particulier, si la notion d'identité se révèlait n'avoir pas de sens en dehors du niveau corpusculaire où elle est associée à l'indiscernabilité totale, ce principe pourrait servir de définition au chaos corpusculaire.

A ce principe s'oppose donc la :

Thèse fondamentale de la Physique : Il existe une Physique, c'est-à-dire une description mathématique du monde sensible, c'est-à-dire que pour ordonner le chaos, on peut postuler la possibilité de relations autres qu'identité.

CHAPITRE 2

INFORMATION ET DURÉE

2. 1 - LE COUPLE DIRECT - INVERSE FONDAMENTAL

2. 2. 1 - AXIOMES DE L'INFORMATION (SCHUTZENBERGER)

Pour pouvoir ramener, comme nous le voulons, les problèmes médians aux problèmes inverses, nous devrons commencer par étudier divers critères possibles d'adaptation. Même en admettant qu'il s'agit de critères d'extremum ou de minimax, le concept de stratégie reste purement verbal, jusqu'à ce que l'on ait déterminé de quelles fonctions de gain ou de risque il s'agit.

Cette détermination se fera à travers un couple direct-inverse fondamental :

- Le problème direct est celui de la détermination de ce qui dans la spécification d'une stratégie donnée est indispensable à l'action, et de ce qui est inutile (par ex. problème d'exhaustivité en statistique).

- Le problème inverse consistera à chercher les formes que serait susceptible de prendre la partie utile des spécifications initiales de la stratégie d'expérimentation, si celle-ci était parfaite, et s'il existait une stratégie d'action parfaite pour en profiter. Ceci est le plus général des problèmes inverses du comportement inductif. Pour reprendre une comparaison de J. Neyman (1949), il est analogue au problème de l'intégrale tel que l'a posé Lebesgue : chercher une fonction de fonction satisfaisant à certaines propriétés. De même, nous chercherons une fonctionnelle de stratégie, que nous appellerons information temporelle, qui satisfera à certaines propriétés. Elle englobera toutes les expressions effectivement construites de parties utiles, tout comme l'intégrale de Lebesgue englobe celle de Riemann, lorsque cette dernière existe.

Elle fournira un principe classificateur de stratégies, d'après le but cherché par rapport à la stratégie d'action. Une fois en possession de la forme fonctionnelle de l'information, nous nous en servirons pour l'étude, par la méthode inverse, de deux théories explicatives : celle de la structure et fonction de la langue, et celle de la structure de la thermodynamique.

2.2 - CLASSES D'ÉQUIVALENCE DE STRATÉGIES DÉFINITION DE L'INFORMATION TEMPORELLE ET DÉTERMINATION DE TOUTES LES FORMES QU'ELLE EST SUSCEPTIBLE DE PRENDRE

Du point de vue mathématique, le problème inverse de la recherche des formes de la partie utile des spécifications a été résolu par M. P. Schutzenberger (1951), qui ne se plaçait d'ailleurs dans aucun contexte physique particulier. Sa solution va être donnée ici presque textuellement.

Nous supposerons que la stratégie est telle que la spécification de ce qu'elle apporte peut être faite avec une seule quantité H, calculable à partir des données sur l'objet, qui sera appelée "information temporelle" et dont les dimensions ne seront pas imposées par les axiomes : ce pourra être un nombre pur, une énergie,etc...

H fera donc partie d'un résumé exhaustif de la stratégie, du point de vue de l'action, et pourra servir de fonction de gain à la stratégie considérée. Elle ne sera vraiment utile que si elle est indépendante des élements indifférents à l'action, donc si le fait d'apporter la même information temporelle introduit une équivalence entre stratégies.

H est rattachée à l'observation qui consiste à déterminer si l'état E_0 pris par E dans une certaine épreuve appartient ou non à l'ensemble X ou $Pr(E \in X) = x$.

Soient X, Y et Z trois ensembles disjoints quelconques de probabilités respectives x, y et z, partitionnant l'ensemble des A_i $(i \in I)$ $(x + y + z = 1)$. Soit CX le complément de X.

La classe d'équivalence de l'observation initiale est définie par les <u>deux conditions suivantes</u> :

(1) qu'il est indifférent de déterminer si $E_0 \in X$ ou $E_0 \in CX$

(2) qu'il est indifférent de déterminer

d'abord si $E_0 \in X$ pour dans le cas contraire su $E_0 \in Y$
ou d'abord si $E_0 \in Y$ puis dans le cas contraire si $E_0 \in X$

Nous voulons que H soit additif pour des stratégies successives et qu'il ait la même valeur pour toutes les stratégies d'une classe d'équivalence. De plus, comme H devra servir à comparer des classes d'équivalence distinctes, il faut ajouter une condition non locale. Il se trouvera que la condition assez faible (1) ci-dessous sera suffisante.

Nous imposerons à H les <u>trois conditions suivantes</u> :

(1) Continuité : H(x) est une fonctionnelle uniformément continue de x.

(2) Symétrie : $H(x) = H(1 - x)$

(3) Commutativité :

$$H(x) + (I-x)H(y/1-x) = H(y)+(1-y)H(x/1-y) = H(x;y) = H(y;x).$$

2.2.2 - FORME FONCTIONNELLE DE L'INFORMATION

Ces axiomes suffiront à donner la forme fonctionnelle de H.

(2) et (3) impliquent que $H(x;y) = H(x;z) = H(y;z)$, et permettent d'associer à toute partition de l'ensemble des A en sous-ensembles disjoints une fonction symétrique de leurs probabilités, qui est la quantité d'information attachée à la détermination de celui d'entre eux contenant E . Schutzenberger pose $K(y;z) = (I-x)H(y/1(x)$. Alors (3) devient :

$$K(x;y+z)+K(y:z) = K(y;z+x)+K(z;x) = K(z;x+y)+K(x;y)$$

dont la solution générale continue est :

$$K(x;y) = f(x) + f(y) - f(x/y).$$

L'homogénéité de degré un de H impliquée par (3) montre que f(x) doit être de la forme x S(log x) où S est un opérateur linéaire quelconque; d'où :

$$H(x) = x\ S(\log x) + (I-x)\quad S\left[(\log\ (I-x)\right].$$

Etant donné le 3ème axiome lui-même, cette formule peut être généralisée au choix entre plus de deux ensembles, par la formule :

$$\boxed{H\ (x_n) = \Sigma_n x_n\ S\ (\log x_n).}$$

2.2.3 - RECHERCHE DE S

Si une stratégie est susceptible de posséder une information, celle-ci devra être de la forme ci-dessus, l'opérateur S étant déterminé par la stratégie. (Cependant la réciproque n'est pas vraie : S ne peut déterminer la classe d'équivalence de stratégie, de toute façon la puissance de l'espace des fonctionnelles linéaires est trop faible pour cela).

Si elle existe, le rôle de l'information temporelle d'une stratégie est de comparer les expériences et de servir de repère du progrès qu'aura dû accomplir la stratégie de détermination pour être en mesure de fournir ce résultat. Un progrès moindre signifiera détermination moins précise, un progrès plus poussé, détermination plus étroite.

Il rend donc inutile l'introduction d'un "temps" extérieur, considéré comme contenant abstrait de l'intervalle entre le début et la fin de la stratégie. Cette notion est remplacée par celle d'information, contenu physique et concret de l'intervalle de temps, intrinsèque aux stratégies d'expérimentation et d'action.

D'où le qualificatif "temporelle". (L'on aurait également pu utiliser le terme "métrique" = "se rapportant à une mesure" - mais il a été utilisé dans un sens différent par Mackay (1950), donc on risquerait confusion) - D'où également le nom d'information que l'on donne intuitivement à toutes les solutions du système fondamental d'axiomes.

Obtenir les expressions possibles de l'information ne résoud cependant nullement notre problème, car on ne peut pas encore identifier les stratégies. La suite naturelle de notre problème inverse serait le problème direct de la détermination de tout ce que l'on peut savoir d'une stratégie en partant de la forme de l'information. Cependant il est de nouveau possible

dans certains cas d'éviter de résoudre ce problème, en le remplaçant par le problème presque inverse de l'identification à des stratégies connues de stratégies optimales correspondant aux expressions inverses S que l'on se donne.

En somme, en cherchant la stratégie - c'est-à-dire la meilleure décomposition fonctionnelle d'un processus donnée par des renseignements globaux - on décide, au lieu de procéder à priori sans utiliser les décompositions familières, de s'appuyer sur des décompositions suggérées par l'expérience des êtres vivants et des machines.

L'élément peu satisfaisant conceptuellement de ce procédé est qu'il ne s'agit pas d'un vrai problème inverse, ces processus "naturels" n'étant à priori optimaux dans aucun sens.

Par ailleurs, deux classes seulement ont été sérieusement étudiées. Elles seront introduites au § 2.3. Leur étude ne fait qu'effleurer le problème de l'expérience. Si la théorie parvient à introduire de nouvelles classes et si celles-ci suggèrent à leur tour de nouvelles méthodes expérimentales, directes, là sera sans doute la principale utilité pratique du niveau d'abstraction auquel nous nous sommes d'emblée placés. Malheureusement, une seule autre classe est connue : indiquée par M. P. Schutzenberger (1951) elle n'a pas encore été étudiée en détail.

Nous ne nous attacherons pas à cette extension,mais montrerons que les stratégies fondamentales de certaines des branches les plus étudiées de la Physique appartiennent à l'une des deux classes fondamentales. Nous verrons ainsi que des lois physiques expérimentales en apparence indépendantes peuvent se résumer par quelques principes universels de l'information, que nous allons énoncer indépendamment d'aucune information particulière. Ce seraient en somme des conséquences nécessaires des processus de mesure, ou en particulier de la perception par les organes des sens (Ceci ne signifiera pas, bien sûr, que le monde s'adapte à ces organes, mais que les organes sont déjà adaptés à la structure physique du monde, étant eux-mêmes des mécanismes physiques).

2. 3 - INFORMATIONS DE FISHER ET SÉLECTIVE

2.3.1 - OPÉRATEURS DE DÉPLACEMENT ET DE DIFFÉRENCE

$$\Delta f(x,\theta) = f(x, \theta + \Delta\theta) - f(x,\theta)$$

$$Ef(x,\theta) = f(x, \theta + \Delta\theta)$$

Posons : $f(x,\theta) = f(x)$, $\quad f(x, \theta + \Delta\theta) = g(x)$

$$H = \int f(x) \log \frac{f(x)}{g(x)}\, dx$$

C'est la seule information faisant intervenir deux lois. Elle s'identifie au résumé exhaustif du sens de Darmois (1936) de la distance entre deux lois de probabilité (Mourier 1946).Développons E en série de puissances :

$$E = 1 + \Delta = 1 + \Delta\theta \frac{\delta}{\delta\theta} + \frac{(\Delta\theta)^2}{2} \frac{\delta^2}{\delta\theta^2}$$

L'information donnée par le 1er terme sera appelée sélective. C'est par ailleurs la seule information concevable pour un message discret.

L'information donnée par le 2ème terme est nulle.

L'information donnée par le 3ème terme sera appelée fishérienne.

2.3.2 - ESTIMATION D'UN PARAMÈTRE ET INFORMATION DE FISHER

Théorème : (Fortet (1950) § 59 et 61) L'inverse du carré de l'écart quadratique moyen d'une estimation d'un paramètre, sans erreur systématique et exhaustive, est égal au nombre de mesures multiplié par l'information de Fisher de chacune des mesures.

Par définition, l'existence d'une estimation exhaustive (qui n'est pas toujours possible) signifiait seulement que l'ensemble des résultats des observations peut être remplacée par une seule fonction de mesures sans aucune perte de spécification. Mais il résulte de plus du théorème ci-dessus que le degré de spécification des résultats peut être mesuré par l'information de Fisher.

Celle-ci est donc effectivement susceptible de jouer le rôle qu'elle ne pouvait jouer que potentiellement du fait de sa définition.

Réciproque du théorème : L'égalité du théorème n'a lieu que si une estimation exhaustive sans erreur systématique est possible.

Définition : on dira que l'on a une échelle propre à distribution propre si une estimation exhaustive sans erreur systématique est possible, qui donne une information fishérienne indépendante du résultat de la mesure lui-même.

2.3.3 - STRATÉGIE SÉQUENTIELLE ET INFORMATION SÉLECTIVE

Dans une stratégie séquentielle, toutes les opérations sont identiques et leur résultat est soit E = arrêter et agir, soit E ($1 \leqslant g \leqslant q$) = continuer. On voit que dans le cas discret, toutes les stratégies sont séquentielles.

Shannon (1948) a abondamment montré que lorsque les opérations sont organisées de façon optimale, leur nombre est proportionnel à ce qu'au § 2.3.1 nous avons appelé information sélective. Celle-ci est donc elle aussi susceptible de jouer effectivement le rôle qu'elle pouvait jouer potentiellement. Le lien de cette information avec la stratégie séquentielle a été mis en évidence par J. Ville et M. P. Schutzenberger (1951) et B. Mandelbrot (1951 b).

Selon la base du logarithme, l'on ajoutera aux termes d'information sélective les qualificatifs népérienne (base e) ou binaire (base 2).

2.3.4 - L'INFORMATION DE POSITION EN GÉNÉRAL

Le concept d'information sélective ne contient bien entendu pas la spécification de l'ensemble de messages d'où est choisi au hasard le message que l'on reçoit.

Un grand nombre de confusions et d'incompréhensions continue de provenir de l'ignorance de ce fait. Elles ne sont pas sans excuse dans les nombreux cas où la distinction entre spécification et message n'est pas claire. En d'autres termes, il n'y a souvent aucun critère pour distinguer entre règles fixes, et objet de gain (Ceci correspondrait à une Constitution incertaine, où la victoire d'un adversaire l'autoriserait à changer la Constitution, et non seulement à voter des lois qui lui sont favorables).

Considérons par exemple le cas d'une série temporelle de signaux. L'instant où le signal se produit peut être attribué :

- à la spécification de l'ensemble d'où est fait le choix. Celui-ci changerait alors avec le temps;

- ou bien au message lui-même, qui comprendrait une intensité plus un temps, c'est-à-dire le N° de l'ensemble d'où l'on fait le choix -

- ou on peut en faire une nouvelle espèce d'information : l'"information de position".

Dans le cas où l'ordre des messages est absolument fixe, et la variable "temps" est une vraie grandeur physique imposée du dehors, le problème est académique, et n'a d'importance que dans la mesure où des confusions en apparence innocentes peuvent conduire à des paradoxes (cf. § 6.4).

Mais cette fixité du "temps" est loin d'être générale dans les messages techniquement importants.

Par exemple en télévision ou facsimile, le "temps" n'est qu'une fausse variable auxiliaire, provenant du procédé de balayage. Le problème de choix ci-dessus représente alors un problème technique important extrêmement réel, car toutes les parties du message attribuées à la spécification deviennent "gelées" du point de vue des possibilités d'adaptation. Il peut être intéressant de les "dégeler", lorsque grâce aux nouvelles possibilités d'adaptation, qui sont ainsi ouvertes, l'on se met en mesure de considérer des propriétés de plus en plus spécifiques du signal à transmettre. L'information de position que l'on doit alors transmettre peut être largement compensée par la possibilité d'obtenir une redondance plus petite dans la transmission de l'information d'intensité.

Dans le balayage uniforme habituel, toute l'information de position est incorporée dans la spécification, c'est-à-dire liée à l'équipement, et l'intensité seule doit être transmise. Ce balayage n'est adapté qu'au message à quadrillage complet, où deux cases voisines diffèrent toujours, et qui en fait ne se rencontrent jamais : donc ce balayage ne permet pas d'utiliser les propriétés réelles de l'image.

Supposons maintenant que l'image ne comporte que des points blancs et des points noirs, les points noirs étant beaucoup moins fréquents que les blancs, ou que les sauts d'intensité soient beaucoup moins fréquents que sur le quadrillage complet. On dispose alors de procédés de balayage à vitesse variable et à retour comme celui de Cherry et Gouriet (1952), qui revient à

intégrer la position et l'intensité en un seul signal (2° méthode).

Supposons que les points noirs au lieu d'être distribués au hasard soient disposés sur des lignes connexes ou non. On peut alors (procédé du pantographe) transmettre de proche en proche la position relative du point suivant : c'est l'exemple extrême de la 3ème méthode, où l'information d'intensité a complètement disparu.

Un exemple moins extrême s'obtient si on s'astreint à traverser les espaces blancs entre courbes par des sauts unité. Il faut alors ajouter un élément d'information d'intensité et l'information de position se réduit à spécifier la direction du déplacement sur le canevas de base. Si on veut que tous les déplacements possibles à partir d'un point aient la même longueur, il faut adopter un canevas en quinconce, et le choix de direction se fait entre six possibilités; l'information par pas est donc au plus de $1 + \log_2 6$.

Mais on peut tenir compte de la rareté des rebroussements, et de la corrélation entre directions des deux pas successifs (corrélation différentielle, intéressante à étudier en soi). Ceci diminue l'information de position et le total doit être de l'ordre de $1 + \log_2 4 = 3$ alternatives par pas, et 3 S alternatives pour une ligne de S points (blancs ou noirs). Le nombre total de points du canevas est bien entendu sans influence sur le nombre d'alternatives à transmettre.

Chacun des deux derniers exemples introduit un problème mathématique préliminaire fondamental : celui de la construction de la "ligne la plus courte" contenant tous les points à transmettre.

2. 4 - PRINCIPES DE LIMITATION DE L'INFORMATION - DURÉES

2.4.1 - INÉGALITÉS FAISANT INTERVENIR L'INFORMATION

Les stratégies optimales sont bien entendu tout-à-fait exceptionnelles, et le concept d'information serait bien limité s'il n'était utile que dans ce domaine.

Mais en fait les égalités des § 2.3.2 et 2.3.3 se généralisent en des inégalités, signifiant que les informations fishérienne, respectivement sélective, restent toujours des limites supérieures de 2 mesures de progrès : respectivement l'inverse du carré de l'écart, et un nombre d'opérations à effectuer.

Les informations, qui sont relatives aux données, et non aux stratégies, continuent donc de constituer des limites à certaines caractéristiques des stratégies. En d'autres termes, si l'on trouve des limites aux informations, concepts d'origine inverse, l'on aura en même temps des limites à ce qu'apporte toute stratégie directe, et qui est mesuré par un "coût" de cette stratégie. En fait, la considération d'un maximum ou d'un minimum relatifs à un problème inverse permet d'oublier la maximation initiale que comportait le passage d'un problème direct à un problème inverse. Les énoncés qui suivent auront donc un caractère inverse par rapport à des problèmes directs c'est-à-dire réels, sans intermédiaire inverse.

Nous énoncerons ces principes de façon un peu détournée. Au lieu de constater qu'entre les éléments possédant la mesure a de la grandeur A, la mesure de la grandeur B possède un minimum b, nous montrons qu'entre tous les éléments possédant la mesure b de B, la mesure de A possède un maximum a.

Les énoncés correspondants exigeront la considération de classes plus larges que les classes d'équivalence du § 2.2. La classe la plus large est bien entendu l'ensemble de toutes les stratégies pouvant s'appliquer à un même objet physique et conduisant à la même fonctionnelle d'information S. On considérera également des "classes restreintes" définies par des critères plus précis.

2.4.2 - PRINCIPES DE LIMITATION

La théorie purement formelle de Schutzenberger du § 2.2 tire son rôle physique et son importance du fait qu'elle fournit un cadre entièrement approprié à l'énoncé des lois physiques, qui sont des solutions, bien entendu d'origine expérimentale et non logique, de certains des problèmes inverses les plus importants de la Physique. Ces solutions s'expriment par des principes du type suivant :

A/ Les contraintes physiques sur l'objet de la stratégie limitent intrinsèquement l'information qu'une stratégie de la classe étudiée est susceptible de fournir entre deux instants donnés. Cette limitation est une nouvelle contrainte physique que ne comportait pas la définition de la classe : elle constitue donc une loi physique.

B/ L'évolution physique de l'objet de la stratégie est telle que si l'évolution indiquée par la règle de comportement inductif est retardée, l'information que l'on possède sur le système ne peut que rester constante ou diminuer de façon non récupérable.

C/ Si les limitations pour les intervalles (1,2) et (2,3) sont égales, celle pour l'intervalle (1,3) est double.

De tels principes sont particulièrement bien connus dans leur application en Quantique. Mais ils sont d'applicabilité beaucoup plus générale, ce qui en justifie l'étude générale. Suivant les termes de P.W. Bridgman (1943), "il y a des situations où on peut montrer par une analyse plus serrée que des dilemmes qui semblent introduits par la mécanique ondulatoire existent déjà en physique classique.

Les faits d'acquérir le maximum d'information entre deux instants ou de le conserver une fois acquis expriment l'adaptation de l'objet et de la stratégie choisie dans la classe considérée, c'est-à-dire une sorte d'équilibre dynamique stationnaire (par exemple : l'entropie se révèlera être une espèce d'information : il y a adaptation si les pertes d'entropie locales dues à certaines opérations sont automatiquement compensées par des gains d'ordre, c'est-à-dire l'entropie zonale ou globale).

On ne peut pas dépasser H mâx, par définition du maximum : une mesure ne peut pas créer d'information. Mais l'inverse est [illegible] Toute stratégie donnant moins que le maximum

est équivalente à une stratégie donnant le maximum suivie d'une perte. Donc $H < H$ max et $\Delta H < 0$ expriment la désadaptation.

2.4.3 - ORIENTATION DE L'INTERVALLE ENTRE DEUX INSTANTS

Chacun des deux énoncés A et B a pour effet d'orienter l'intervalle entre deux instants. Toutes ces orientations sont identiques. En effet, pour le premier, si deux stratégies établissaient des orientations opposées, on pourrait les accoupler dans un cycle et chacune des deux informations pourrait croître indéfiniment, ce qui contredirait la loi de limitation de l'information. Pour le deuxième énoncé, on pourrait couplet l'évolution passive avec une stratégie active, ce qui permettrait de récupérer l'information perdue, ce qui contredit la nature des pertes.

2.4.4 - CONCEPT DE DURÉE

Les principes du type B, et en particulier un des aspects du deuxième principe de la Thermodynamique, ne permettent qu'une orientation des intervalles de temps, et ne peuvent pas fournir une structure du temps. D'où l'insuccès habituel des tentatives pour relier le temps à l'information. Mais cet insuccès ne tient pas à un vice du concept d'information, mais à un mauvais choix du principe. En effet, les principes A permettent d'inverser toute loi disant que le maximum d'information est proportionnel à la durée d'expérimentation. Ce faisant, nous obtenons une définition du type suivant :

La durée entre deux instants, propre à une classe de stratégie, est le maximum de l'information que peut apporter une telle stratégie entre ces instants.

Si une stratégie peut être incorporée dans une classe, sa durée propre sera l'information qu'elle apporte.

La notion de durée n'est ni extensive, ni intensive. Elle est strictement locale (dans un sens plus strict même que celui de la Relativité, car il n'exige pas la notion d'espace, qui peut donc être déduite de celle de durée. Nous retrouverons là un procédé de E.A. Milne (1948), mais on n'a plus à utiliser le "contenu" métaphysique que constitue la "conscience" : on peut se baser sur un contenu physique concret). (Remarquons cependant que pour qu'une expérience de "durée" Δt donne effectivement le maximum d'information, il peut être nécessaire qu'elle prenne théoriquement un temps infini; mais elle reste finie dans ce sens qu'on peut recommencer une autre expérience, sans beaucoup perturber la première).

En somme, conformément à l'exemple d'Einstein définissant l'espace par les masses qu'il contient, nous ne considérons pas la durée comme un contenant abstrait; c'est ce que mesure une horloge quelconque. C'est cela, et rien de plus.

Chacune de ces horloges n'aura pas à être considérée comme tentative (imparfaite) de saisir une "idée transcendante". Au contraire, tout appareil, tout mode opératoire, auront leur durée, qui sera aussi bonne qu'une autre, puisque c'est dans son cadre que devra être faite la théorie de cette horloge : En d'autres termes, l'information peut à posteriori servir de para-

mètre fondamental au demi-groupe de transformations de l'objet que constituent les opérations d'une classe d'équivalence de stratégies : les paramètres de ces opérations ont été déduits de leurs propriétés, et leur sont particulièrement intrinsèques.

Nous verrons que le cas fondamental où les principes ci-dessus sont vérifiés est celui de la durée thermique, qui mesure le maximum d'information fishérienne que l'on peut extraire sur l'amplitude d'un signal, ainsi que la durée sélective, qui se rattache à la durée thermique.

La limitation de l'information confirme l'unicité de direction du temps qui résulte de l'intuition et qui est interprétée par l'accord et l'équivalence des temps de toutes les fonctions organiques. De plus, les limitations d'information étant liées au temps perdent toute idée d'imperfection et prennent un caractère irréductible.

2.4.5 - LIMITES DE STRATÉGIES

Le passage à la limite qu'exige la définition de la durée peut comporter des difficultés du point de vue de la théorie mathématique.

En effet, une stratégie réelle est toujours une suite de mouvements discrets se succédant dans le temps (et définissant la durée). Ce caractère discret correspond à l'impossibilité d'une proposition logique infinitésimale, aspect de la quantification qu'entraîne nécessairement la logique propositionnelle.

Cependant, le caractère discret n'intervient pas dans les axiomes de l'information, qui resteraient valables si on passe à des mouvements de plus en plus courts et nombreux. Mais on passe alors à une stratégie-limite et non une logique-limite non propositionnelle qui n'appartiennent pas à la théorie de von Neumann. Tel sera le cas pour la capacité informationnelle sélective potentielle qui sera définie au § 4.3.

Le seul cas où des stratégies-limites ne se présentent pas est celui où le maximum d'informations correspond à une vraie stratégie unique. La durée correspondante resterait alors quantifiée comme elle l'était dans les stratégies non maximales. Le quantum = durée d'une expérience élémentaire est alors irréductible, inanalysable, et par suite incompréhensible (cf. J.R. Oppenheimer : "Une particule élémentaire c'est quelque chose de tellement simple qu'on n'y comprend rien du tout"). La durée serait alors le maximum du nombre des quanta ou des mouvements (expériences élémentaires) entre deux instants.

Ce cas correspond à l'existence physique de "propositions élémentaires".

Il peut bien entendu arriver que le maximum d'information corresponde à une vraie stratégie unique simplement parce que la classe considérée est restreinte. Par exemple, c'est le cas, par rapport à la classe séquentielle, si on restreint la largeur de bande W d'une ligne de transmission. Dans ce cas, le nombre de "quanta" par seconde aura une grande importance, bien qu'il soit sans liaison proportionnelle avec l'information temporelle transmise. Ce nombre a été appelé "information" structurale par Gabor et Mackay (1950). Mais il n'a ni les propriétés ni les di-

mensions d'une information : nous appellerons 2W capacité structurale. On peut grouper sous le terme d'information structurale de la stratégie tout ce qui permet de spécifier le caractère de son résultat avant qu'il soit connu et qui n'entre pas dans l'information temporelle : la capacité structurale en ferait donc partie.

Un autre cas où les stratégies limites se présentent mais qui ne donne pas lieu à difficulté est celui où l'information est la même pour toutes les stratégies vraies considérées : si le nombre d'opérations double, l'information par opération diminue de moitié. Alors il n'y a pas de quantum de temps intrinsèque, ni "propositions élémentaires" (c'est pourtant dans un tel cas que Mackay (1950) a voulu baser sur celles-ci une théorie de l'information. Cette base est donc fort incertaine).

On se trouve très souvent très près du cas ci-dessus du maximum entièrement plat, car en pratique il est facile d'arriver tout près de la perfection à l'aide de stratégies très simples tandis que les derniers progrès coûtent beaucoup en complication (cf. Fisher : le gain qui peut être réalisé par des améliorations de détail des techniques statistiques ordinaires est très faible). (Donc l'existence du temps instinctif des organes n'exige pas que leurs stratégies soient parfaites).

Notre point de vue général entraîne la possibilité d'autres cas que ceux qui précèdent. Il englobe aussi la possibilité de stratégies hiérarchisées, un mouvement du niveau supérieur étant une stratégie du niveau inférieur, et correspondant à une "proposition élémentaire" indécomposable dans le niveau supérieur.

CHAPITRE 3

DÉMONS DE MAXWELL

3.1 - COMPARAISON DES DURÉES

3.1.1 - DURÉE COSMIQUE

Chacune des notions de durée, prise séparément, n'exige pas la notion de temps; il n'en demeure pas moins que l'intérêt de ces notions vient de ce que l'on reconnaît à posteriori l'équivalence des durées comme repère du progrès des stratégies. Ceci traduit en quelque sorte la non-contradiction entre les diverses disciplines de la Physique, et s'écrit en inversant l'une des nombreuses lois physiques exprimées d'habitude par la proportionnalité entre une information et le "temps" de nature cosmique "dans" lequel se passent tous les phénomènes.

Celui-ci étant sans liaison à priori avec les phénomènes étudiés, notre interprétation par "durée proportionnelle au temps cosmique" paraît préférable à la "loi" inverse, car elle part du mieux défini, plus local et plus fondamental des deux termes de l'égalité.

L'introduction d'une durée cosmique par comparaison des durées peut se faire expérimentalement par étapes successives, qui introduiront ainsi, au fur et à mesure que le champ de la comparaison s'élargit, une série de durées intermédiaires de plus en plus intrinsèques. Si l'on peut considérer "toutes" les stratégies, il existera une limite à la comparaison de leur durée.

Capacités.

Si l'existence d'un temps physique commun "public" est admise, la définition de la durée doit être suivie d'une définition du type suivant servant à comparer les diverses durées au temps commun:

Définition.

La capacité informationnelle d'un ensemble de stratégies est le maximum de l'information qu'elles sont susceptibles de donner par unité de temps (en particulier la capacité d'une stratégie d'un type est l'information qu'elle transmet par seconde).

La capacité d'une suite de stratégies est la plus petite des capacités des stratégies de la suite. L'identité des capacités sera une condition nécessaire d'adaptation.

Il est souvent intéressant de considérer des durées intermédiaires pour elles-mêmes : par exemple, le temps "privé" d'un mécanisme résulterait de ses diverses stratégies comme optimum, moyenne, ou durée de la stratégie prédominante à un instant donné.

Un autre exemple est celui des méthodes visuelle et orale de détection en Radar (Cf. Lawson-Uhlenbeck, 1950, p. 165). Lorsque chacune est poussée à la limite, elles sont essentiellement équivalentes, fournissant la même quantité d'information. Mais la limite absolue, due au bruit thermique, n'est pas atteinte par ces méthodes.

Un procédé général de comparaison de stratégies a été utilisé par Gabor (1950). Il élimine progressivement toutes les particularités de l'appareil pour arriver à une information - donc un temps - plus intrinsèque.

Procédé inverse. Les axiomes du temps et de l'information étant identiques, la proportionnalité au temps de toute expression spécifiant la connaissance d'un objet permet de conclure qu'elle appartient aux informations temporelles dont elle remplit tous les axiomes, même si l'on est incapable d'écrire l'opérateur S et d'identifier la stratégie correspondante.

3.1.2 - DÉMONSTRATION D'ÉQUIVALENCE

Une démonstration exige ou bien un modèle du type cinétique ou mécanique statistique, ou bien un mécanisme qui évite ces modèles en ramenant l'équivalence à des principes plus simples, de caractère phénoménologique. Nous appellerons ces mécanismes les démons de Maxwell.

Les démons s'introduiront ainsi dans cuacune des étapes de la comparaison entre stratégies. Les premières étapes donnent les notions d'information et de durée propre; les suivantes comparent l'utilisation d'un même intervalle par deux stratégies de S différents.

3.2 - DÉMONS DE MAXWELL

3.2.1 - LE DÉMON ORIGINAL ET SES RÉFUTATIONS

On peut lire à la p. 328 de la "Theory of Heat" de J. Clerk Maxwell (1871) :

" Soit un être dont les sens sont si fins qu'il peut suivre
" chaque molécule dans son mouvement et effectuer des opérations
" qui nous sont à présent interdites. Soit un récipient divisé
" en deux sections A et B par une paroi percée d'un petit trou.
" Supposons qu'un être qui peut voir les molécules individuelles
" ouvre et ferme ce trou, de façon à ne laisser que les molécu-
" les rapides passer de A en B, et les lentes de B en A. Sans
" travail, il élevera le température de B et abaissera celle de
" A, en contradiction avec la 2ème loi de la Thermodynamique."

Cette conclusion paradoxale était bien entendu la raison pour laquelle le démon avait été imaginé. Nous allons étudier ici les diverses méthodes de réfutation. Nous n'en conclurons

pas cependant que le "démon de Maxwell ne peut pas exister", mais que "le démon de Maxwell ne peut avoir un rendement supérieur à 1". Cette dernière forme nous permettra de conserver un nom commode et consacré à des instruments dont l'importance théorique est très grande dans le contexte présent.

Toutes les réfutations, très diverses, de la conclusion de Maxwell, reposent sur un même principe : on attribue une réalité physique à une, et une seule, des opérations que le démon doit effectuer. Les autres opérations peuvent être laissées dans l'ombre initiale, car une seule opération suffit à la démonstration, et que considérer toutes les opérations aurait dépassé les possibilités du raisonnement.

Remarquons que l'erreur de Maxwell est tout-à-fait compréhensible, car la considération de cette opération était inutile dans la Physique à échelle humaine; elle n'est devenue indispensable en pratique que dans un problème bien récent : celui de la transmission de l'information. Dans ce cas, une quantité d'habitude finie était devenue infiniment petite. Il était par suite devenu indispensable d'introduire dans le bilan où elle entrait, toutes sortes de quantités qui n'avaient jamais cessé d'être infiniment petites : les bilans des "sens" du Démon.

Il suffit de choisir une seule d'entre celles-ci pour que le bilan cesse d'être absurde. D'ailleurs, le choix de l'opération, fondamental dans la réfutation du paradoxe de Maxwell, reste en grande partie arbitraire. Il en résulte que les diverses démonstrations d'impossibilité s'appliquent en réalité à des êtres tout-à-fait différents (certains seront étudiés au § 4.5).

3.2.2 - DÉFINITION DU DÉMON

Elle se fait en deux étapes.

Tout d'abord, par définition de la durée, l'information fournie permet de comparer entre elles une stratégie utilisée et une stratégie repère, si elles font partie d'un même ensemble, où la 2ème est optimale du point de vue de l'information commune.

Nous appellerons démon de Maxwell tout instrument utilisant une stratégie non optimale.

Il se trouvera que la stratégie repère des démons de Maxwell de la Thermodynamique est une stratégie minimax (réversible, isostatique, cf. Chap. 4), qui par définition transmet mieux l'information sélective que toute autre stratégie. Donc la définition précédente suffit à englober beaucoup de démons classiques.

Mais pour les englober tous, il faut généraliser la définition, et ceci fera également entrer des instruments que l'on n'a pas l'habitude d'appeler démons, et en particulier tous les procédés techniques imaginables de modulation.

Dans la définition la plus extrême, on appellerait démon de Maxwell, relativement à une définition donnée de l'information et une stratégie repère transportant l'information considérée, tout instrument utilisant une autre stratégie quelconque.

Les démons ainsi définis seraient des cycles fermés composés de deux demi-cycles réunissant les états initial et final et parcourus en sens inverses.

Cependant, un démon aussi général risquerait de pouvoir avoir un rendement supérieur à 1. Si l'on veut conserver la ressemblance avec le démon initial, il faudrait :

- ou bien ne prendre la stratégie utilisée que parmi celles qui sont moins favorables que la stratégie repère,
- ou bien ne prendre comme repère que des stratégies absolument optimales par rapport à leur information propre.

L'une et l'autre de ces deux définitions donnerait une claire conception de la multiplicité des démons, au moins aussi nombreux que les informations.

Ces démons pourront alors être utilisés comme "expériences fondamentales" de branches entières de la physique, et ceci de deux façons distinctes, correspondant aux deux définitions :

A/ Les deux stratégies sont considérées symétriquement, les propriétés de l'une et l'autre étant connues. Alors :

- ou bien on démontre l'impossibilité d'un démon ayant un rendement supérieur à 1 (cette démonstration est transitive en ce sens que si (1,2) et (2,3) sont impossibles, (1,3) l'est aussi automatiquement : avec 3 stratégies; il suffit donc de 2 démonstrations.)

- ou bien on constate qu'un tel démon est possible; alors on est assuré que la stratégie repère n'est pas intrinsèque pour l'information de base.

B/ Les stratégies sont considérées asymétriquement, certaines de propriétés de celles qu'on utilise étant inconnues, ou, à la limite, la stratégie utilisée étant plus ou moins indéterminée.

- Supposons que l'on puisse poser un principe d'impossibilité de rendement > 1. Nous verrons que le principe de Carnot est de ce type, relativement à une stratégie minimax et à l'information sélective. A priori, il pourrait y avoir d'autres tels principes. Sinon, la possibilité de ce principe donnerait un caractère particulièrement intrinsèque à la stratégie et à l'information auxquelles il s'appliquerait; et sa durée repère devrait être prise comme durée fondamentale. Réciproquement d'ailleurs, le caractère absolu d'une durée entraîne l'existence d'un "principe de Carnot".

Un principe d'impossibilité entraîne d'autre part comme conséquence des propriétés nécessaires de toutes stratégies utilisées. Si ces propriétés menaient à des contradictions, le principe d'impossibilité ne pourrait être vrai. Si ces propriétés sont vraies, il ne faut pas en déduire des analogies factices entre les stratégies utilisées.

Le démon réversible-quantique a été étudié suivant le 1er procédé par P. Demers (1944) et L. Brillouin (1951, Cf. § 4.5.3) par le 2ème procédé par P. Demers (1945) et D. Gabor (1951) qui ont essayé d'en déduire le principe des quanta ou la relation d'incertitude de Heisenberg.

3.2.4 - EXEMPLES DE DÉMONS

Les démons les plus importants restent ceux dont la stratégie de repère est la stratégie isostatique. Leur étude doit donc être remise au Chapitre suivant : elle présente un intérêt fondamental pour fonder la Thermodynamique pure, sans référence aux quanta ou à d'autres phénomènes exigeant des modèles.

DEUXIÈME PARTIE

CHAPITRE 4

THERMODYNAMIQUE DU SIGNAL PARFAIT

4.1 - DURÉE THERMIQUE ET TEMPÉRATURE

4.1.1 - BRUIT THERMIQUE

Soit un générateur de signaux. D'après le théorème de Thévenin, il peut être représenté par une force électromagnétique E en série avec une impédance R + jX. Le maximum de puissance qu'on peut en tirer est $S = E^2/4R$ qu'on appelle la puissance disponible.

La puissance disponible ne peut jamais être mesurée avec exactitude. Tout se passe comme s'il s'y ajoutait une source de signaux aléatoires de Laplace-Gauss de puissance donnée par la loi de Nyquist (le concept de puissance est directement lié à ceux d'intensité et de résistance et n'exige pas celui de temps)

$$\delta = \frac{E_b^2}{4R} = \frac{4R\ kT\ W}{4R} = kTW = \frac{kT}{2\tau}$$

où T est la température Kelvin et W la largeur de bande du signal, donc $\tau = 1/2W$ le temps disponible pour chaque mesure : ce temps sera pris comme mesure de l'indétermination sur l'"instant précis" de la mesure.

La loi de Nyquist est susceptible d'une interprétation informationnelle fishérienne. En effet, $1/\delta$ est proportionnelle au carré de l'erreur probable sur l'amplitude du signal, donc c'est l'information fishérienne sur l'amplitude.

Multiplions-la par le nombre de mesures par seconde - qu'on a suggéré d'appeler capacité structurale du signal, (§ 2.4.5). Le produit $(1/\tau)$ $(1/\delta)$; que l'on peut appeler information fishérienne spécifique, est une constante. La loi de Nyquist peut dès lors s'énoncer sous une forme analogue à celle du principe zéro de Guggenheim (1949) qui définissait la température :

L'information fishérienne spécifique sur l'amplitude d'un signal ne dépend que d'un seul paramètre caractérisant le signal et non pas du rythme et des modalités de la mesure (c'est-à-dire du nombre $1/\tau$).

4.1.2 - DURÉE THERMIQUE

Cette loi a exactement la forme qu'il faut pour être interprétée par une durée, en dehors de tout bruit. Ceci est d'ailleurs logique car si les instruments disponibles sont tous à l'échelle thermique, le bruit ne peut pas, donc ne doit pas, être interprété comme un signal parasite imprévisible se superposant au signal prévisible.

La forme de la loi de Nyquist étant indépendante de toute géométrie, on peut d'abord prendre $1/\delta$ comme définition de la durée thermique.

Ensuite, on compare les durées à diverses conditions physiques : on constate qu'elles sont équivalentes (proportionnelles); il suffit donc de les pondérer par un seul paramètre qui caractérise les propriétés de l'état physique du signal et l'unité de durée. Nous l'appellerons température f, kT_f (f signifiant fishérienne ou de fluctuation suivant le point de vue où l'on se place).

kT_f est un coefficient de similitude des durées thermiques, mais, en fait, sa grandeur importe peu tant qu'on ne considère que des opérations à température constante : seule importe son existence et sa non-nullité, qui exprime la limitation de l'information fishérienne.

La comparaison effective des T_f ne peut se faire que sur la base d'une durée intrinsèque commune. Comme telle nous adopterons au § 4.4.4 la durée sélective, et les cycles de Carnot permettront dès lors de déterminer les rapports entre températures de fluctuation, c'est-à-dire entre fluctuations à températures différentes. L'unique coefficient arbitraire qui subsistera proviendra de l'indétermination qui subsiste pour l'échelle unique de durée sélective.

La température est une grandeur physique, si l'on convient que cette notion (distincte de celle d'objet physique) s'obtient de façon générale en effectuant le quotient (au sens de la Théorie des Ensembles) de l'ensemble des objets par une relation d'équivalence. (Cette notion de quotient est une généralisation de la définition de l'"axe des x" comme "quotient" du plan xy par la relation d'équivalence : deux points sont équivalents s'ils ont même abscisse). Dans le cas présent, l'équivalence est constituée par l'égalité de deux énergies intensives. Au § 4.3, on aura des équivalences constituées par l'égalité de nombres purs, C, B, S, fonctions de l'état du signal.

Remarquons que cette définition de la température et en particulier l'indépendance de la géométrie, a en fait exigé plus que le principe de limitation général posé au Chapitre 2. Ce principe seul aurait exigé l'introduction d'un coefficient numérique fonction de la géométrie, que l'on peut présumer être très simple, mais que l'on suppose ici égal à 1.

4.1.3 - CARACTÈRE DE LA STRATÉGIE DE LA NATURE

Nous n'allons pas dans ce qui suit essayer d'analyser les "raisons" de l'existence d'une température, en la ramenant à d'autres phénomènes. Par suite, notre point de vue stratégique et informationnel en Thermodynamique ne sera nullement explica-

tif. C'est un point de vue inverse, qui a pour but de préciser, relativement à l'observateur, certaines limitations phénoménologiques, en les réinterprétant comme aspects de la "durée thermique".

(Le fait qu'il s'agira uniquement de théorie non-quantique conduira à prendre des limites qui en fait auraient fait entrer dans le domaine quantique. Mais il se trouve qu'on peut construire une théorie non quantique conceptuellement très homogène, et ensuite seulement limiter son domaine de validité et introduire des corrections quantiques dans la zone de transition).

Il semble cependant opportun de discuter le caractère de la "stratégie de défense" de la Nature du point de vue de la continuation possible de la théorie présente par une explication de la durée thermique en termes plus élémentaires. Cette stratégie pourrait être aussi bien :

- pure : interprétation déterministe; les circonstances de chaque mouvement le déterminent parfaitement, mais il n'existe aucune contre-stratégie : le bruit serait la partie du message pour laquelle il n'existe aucune contre-stratégie (au niveau considéré) (par suite, si la contre-stratégie du signal, réalisable, n'est pas effectivement réalisée, la partie incompréhensible du signal devient indiscernable du bruit).

- entièrement aléatoire : le bruit serait "vraiment" du bruit : interprétation indéterministe et hypothèse d'absence de variables cachées;

- ou mixte, les circonstances ne déterminent que les probabilités relatives de divers mouvements.

La troisième alternative englobe la première et la deuxième; elle est donc plus riche. Elle permet aussi d'envisager le cas où l'indétermination intrinsèque serait répartie entre deux causes : 1° les limitations temporelles, 2° l'imperfection du décodage. C'est l'interprétation en fonction de variables cachées indéterministes.

Si l'on ne sort pas d'un noveau donné d'analyse, il n'y a aucun moyen de décider laquelle des interprétations est la bonne, ni même aucun intérêt à le faire. Mais on devient en mesure de décider si l'on introduit des modèles, tels que la Théorie Cinétique. Celle-ci interprète la stratégie thermique comme mixte; les limitations d'information associées au temps thermique se ramenant alors dans le cas classique à des conséquences d'une stratégie pure basée sur le temps mécanique associé à la vitesse, et dans le cas quantique, aux conséquences d'une stratégie mixte associée au temps quantique.

Par sa définition même, le temps sélectif thermique n'a de valeur intrinsèque que dans la mesure où il existe une Thermodynamique indépendante de la Mécanique Statistique : nous voyons ici qu'il n'a pas de caractère intrinsèque par rapport au niveau inférieur.

4.1.4 - TEMPÉRATURE ET ÉQUILIBRE

D'habitude, la température n'est définie qu'après qu'une notion d'équilibre ait été introduite (principe zéro de la Ther-

modynamique selon Guggenheim (1949). Voir aussi Landé (1926) p. 284). Ici cette étape n'a pas été explicitée. Mais en fait dans la définition même de l'information fishérienne étant contenu un concept d'"erreur probable". Celui-ci implique la considération d'un "ensemble d'objets" identiques, du point de vue de l'instrument d'observation, au signal étudié, qui est de dimensions "microscopiques". Nous pouvons considérer que ce petit signal est en équilibre avec des signaux identiques. On retrouve là un procédé fréquent en Mécanique Statistique, mais utilisé en général à plus grande échelle. Ici la température donne précisément la limite des objets dont l'intérieur peut être considéré comme étant en équilibre.

Einstein aurait fait remarquer à propos du démon de Maxwell que "dans un milieu en équilibre, il semble probable qu'un mécanisme intellectuel ne pourrait agir". Un signal est un mécanisme intellectuel et la température est par définition la dimension en deçà de laquelle il cesse de pouvoir fonctionner.

Le concept de particule comme celui de quantum reste en dehors de la Thermodynamique telle que nous la concevons ici. Cependant, ce qui précède peut aider à l'introduire en quelque sorte axiomatiquement. A une température donnée, kT donne la dimension moyenne en énergie des signaux non décomposables : atomiques. C'est bien entendu conforme au théorème et principe d'équipartition de l'énergie.

4. 2 - INFORMATION SÉLECTIVE ET STRATÉGIES SÉQUENTIELLE ET SÉLECTIVE

4.2.1 - INFORMATION SELECTIVE AVEC BRUIT

S'il n'y a pas de bruit, on peut considérer la Nature comme inactive; la transmission peut se faire de façon discrète et l'information sélective est la seule à considérer, et ceci sous la forme discrète, car elle constitue la seule fonction de gain appropriée au récepteur.

Il n'en est plus ainsi lorsque la Nature agit en perturbant le signal reçu par l'observateur, et qu'il y a vrai duel.

Il faut insister, avec P.M. Woodward (1951) sur ce caractère nécessairement continu et souillé de bruit de tout signal réel. La question est de savoir si on introduit de l'information apparente ou si l'on perd de l'information réelle en remplaçant le nombre infini de valeurs de ce signal continu par un nombre fini de décisions. C'est un problème d'exhaustivité, qui n'a pas été touché par Woodward. En fait, pour simplifier, nous supposerons dès maintenant que le paramètre étudié est susceptible d'estimation exhaustive, donc que le progrès de la stratégie d'estimation peut être suivi aussi bien par l'estimation elle-même (Cf. § 2.3.2) que par l'information fishérienne (celle-ci étant égale et non pas supérieure, au carré de l'inverse de la variance). Il se trouvera d'ailleurs que cette hypothèse n'est pas restrictive, car la distribution du paramètre étudié dépend de la stratégie, et tous les signaux optimaux devront être de Laplace-Gauss, c'est-à-dire susceptibles d'estimation exhaustive.

Etant donné l'existence du bruit, il faut, en plus de l'information apparente contenue dans le signal reçu, considérer une fonction de risque. Shannon a montré qu'il existe une fonction de risque, qu'il appelle équivocation, telle que la différence entre l'information apparente et cette fonction de risque représente la limite de l'information réellement recevable lorsque le pourcentage d'erreurs tend vers zéro. Mais ni l'information apparente, ni l'équivocation, ni leur différence ne sont elles-mêmes des informations (Cf. Blundell 1952). (Les erreurs les plus absurdes peuvent résulter de la confusion entre gain et information.).

Pour arriver à ce résultat, Shannon pose d'abord le problème de savoir si l'information apparente transmissible par un signal donné ne tend pas vers 0 avec les erreurs maxima de décodage. Toute action inductive exacte serait dans ce cas impossible, ou exigerait un signal infiniment long ou infiniment fort. Autrement dit, pour "vaincre l'opposition de la Nature", il faudrait utiliser des ressources en temps et en énergie infinies. Ce problème de Shannon est par suite tout-à-fait fondamental pour la théorie du comportement inductif en Physique. La solution que lui a donné Shannon souffre de certains défauts sérieux, mais il n'en reste pas moins qu'on peut l'utiliser dans beaucoup de domaines : elle est en particulier fondamentale pour le problème de la réversibilité, dont on montrera plus loin qu'il constitue un commentaire de ce théorème de Shannon.

La solution de Shannon considère une stratégie qui réalise une information apparente et une erreur données, et le signal qui représente cette stratégie. On considère ensuite toutes les méthodes de lui associer d'autres signaux, c'est-à-dire d'autres stratégies. On montre que $\sqrt{\varepsilon}$ est une limite supérieure du pourcentage de stratégies pour lesquelles la probabilité d'erreur peut dépasser $\sqrt{\varepsilon}$. ε peut être rendu arbitrairement petit lorsque le délai peut augmenter indéfiniment. Quant à l'information sans erreur portée par les stratégies les plus nombreuses, elle s'obtient en supposant donnée la ligne, c'est-à-dire la stratégie de la Nature; et prenant le maximum par rapport à toutes les stratégies de l'observateur de la différence entre l'information apparente et une fonction des erreurs appelée équivocation. Malheureusement les stratégies à faible erreur, bien que de loin les plus nombreuses, sont pratiquement impossibles à construire.

Il est en somme possible de détruire l'interaction en apparence indestructible entre les deux joueurs en remplaçant la Nature par un joueur sans résistance, mais en revanche l'observateur par un joueur obtenant un gain moindre que le gain apparent.

Nous appellerons stratégies sélectives les limites ainsi introduites des stratégies séquentielles (ces limites ne sont peut-être plus de vraies stratégies au sens de von Neumann, Cf. § 2.4.5). Le gain correspondant est la "valeur" du jeu pour l'observateur, à Nature constante; c'est une information potentielle.

4.2.2 - INFORMATION COMME FONCTION DE GAIN. MAXIMUM. MINIMAX

Dans ce qui précède, le bruit était gaussien. Mais tout ce qui importait, c'est que la stratégie de la Nature soit fixée : la théorie aurait pu s'appliquer à toute autre forme du bruit.

La fonction de gain dépend de la stratégie de l'observateur. Si ses actions sont rendues encore plus libres, à savoir si la forme de la distribution des valeurs du signal peut être modifiée, à puissance moyenne fixée, le maximum d'information s'obtient pour un signal normal de Laplace-Gauss.

Mais de plus la fonction de risque dépend du bruit de la même façon fonctionnelle : le gain est la différence entre les valeurs que prend une même fonction pour deux variables : les deux joueurs. Donc la stratégie du signal qui maximise le gain est identique à la stratégie de la nature qui le minimise. Par suite, il se trouve que la forme gaussienne que prend le 2ème terme du gain est celle-là même qui aurait correspondu au cas où la Nature aurait cherché à diminuer la valeur du jeu. Donc l'information potentielle, ou capacité d'un signal souillé de bruit, est également le minimax ou maximin de la fonction de risque (ici l'ordre importe peu, car les deux parties sont indépendantes).

Tout se passe donc comme si, dans ce premier exemple où le critère du jeu doive être précisé, ce critère se révélait être précisément le critère minimax original de J. von Neumann et A. Wald.

C'est là pour la stratégique un important succès, qui se confirmera au § 4.3.5.

Il y a donc identité de structure entre le bruit le plus défavorable et le signal le plus favorable : le signal le plus incompréhensible si on ne connaît pas le code est précisément celui qui aurait transmis le plus d'information si on connaissait le code. Par suite, le signal qui sur une ligne donnée transmet le plus d'information est celui qui a la structure statistique du bruit, indécodable, propre à cette ligne. C'est là un résultat valable aussi bien pour des lignes analytiques qu'arithmétiques. Il ne veut toutefois pas dire que le "bruit soit ce qui transmet le plus d'information" : pour transmettre de l'information, il faut être deux : émission et réception. Le bruit vrai est ce pour quoi il n'y a pas de réception possible : c'est aussi l'impression que donne tout signal parfait si on veut le décoder, de travers, ou toute grandeur physique lorsqu'on veut la mesurer à l'aide d'un appareil (par ex. circuit) mal adapté.

Une corrélation entre le bruit et le signal rendrait le risque différent d'une simple différence. Les actions de l'adversaire étant un peu liées à celles de l'observateur, l'équivocation est moindre. On a affaire à une coalition qui augmente l'information potentielle.

De même, si la stratégie de la Nature était autre que gaussienne (mais susceptible d'estimation exhaustive) l'on ne serait plus au minimax, et l'information qu'elle laisse acquérir serait supérieure, à puissance de bruit donnée.

4. 3 - MESSAGES CONCRETS ET TRANSFORMATIONS ISOSTATIQUES ENTROPIE

4. 3. 1 - DEFINITIONS

Le signal que nous avons considéré aux § 4.1 et 4.2 était une valeur "instantanée". Considérons maintenant non plus une valeur seule, mais l'ensemble des valeurs successives. Ceci est réaliste, car si l'on néglige, comme nous le faisons, la structure particulaire de l'énergie et de la matière, l'information sélective du message est transmise, en dernière analyse, par les variations d'un signal en fonction de sa durée propre (qui est la durée thermique : par suite, la durée sélective sera à priori proportionnelle à la durée thermique à partir de laquelle elle est construite).

Les propriétés du signal, du point de vue de la transmission de l'information, doivent en principe être déduites de son "état physique". Mais ceci aurait exigé de préciser exagérément le genre de signal dont il s'agit. Pour l'éviter, nous allons encore une fois procéder inversement : nous supposerons donnée une mesure pré-thermodynamique de signal : masse ou voltage, et définirons le signal par ses propriétés du point de vue des deux joueurs : Expérimentateur et Nature, auxquels correspondent deux informations spécifiques différentes. Il se trouvera que la température devra être complétée par un paramètre spécifiant l'extension des domaines de stratégies. La description numérique de l'état comportera donc au moins deux variables d'état (Deux sera aussi le nombre minimum de variables pour toute autre description rendue possible par la connaissance de l'"état quantitatif" du signal : gaz parfait, etc...). L'identité d'état entre deux signaux s'exprimera par deux égalités.

Un signal ainsi définissable par deux variables sera dit parfait.

Supposons que les variations du signal soient entièrement utilisées, c'est-à-dire que toute mesure indépendante ait une signification informationnelle. Nous dirons alors que le <u>message que porte le signal est concret</u>. <u>La stratégie</u> qui fait correspondre un tel message avec le signal <u>sera dite isostatique</u>, de même que la transformation que subit le singal lorsque l'on remplace la connaissance de toutes ses propriétés macroscopiques par la connaissance du message qu'il porte.

Pour introduire les fonctions d'état autres que la température, nous calculerons d'abord le maximum d'information sélective, en l'absence de toute contrainte autre que celle que traduit la température. Nous mesurerons ensuite ces autres contraintes par la diminution de l'information maximum qu'elles provoquent et qui sera appelée <u>entropie absolue</u>.

4. 3. 2 - SIGNAUX A ENTROPIE NULLE (PROCESSUS DEGENERES)

Ce seront ceux dont l'état permet de réaliser le maximum d'information sélective à température donnée. Leur énergie sera dite entièrement libre.

Soit T la température du signal. Elle introduit entre les indéterminations de durée thermique τ et de puissance disponible δ la relation $\tau\delta > kT/2$ (§ 4.1).

D'après les propriétés des stratégies sélectives, si τ, δ et la puissance moyenne P du signal sont donnés, le minimax de l'information sélective népérienne est, par définition de C' et de la fonction G :

$$B' = \frac{1}{2\tau} \log_e \left(1 + \frac{P}{\delta}\right) = \frac{P}{kT} \left[\frac{\delta}{P} \log_e \left(1 + \frac{P}{\delta}\right)\right] = C'G \left(\frac{P}{\delta}\right) = \frac{1}{2\tau} \log_e (1 + 2\tau C')$$

Supposons maintenant τ et δ variables, et poursuivons notre recherche du minimax, qui se réduit à un maximum, car seul l'observateur élargit encore son champ de stratégies. B' croît lorsque τ décroît : on a donc avantage à confier l'information à des mesures aussi imprécises et fréquentes que possible. Le processus stochastique que constitue ce signal est "dégénéré de deuxième espèce" dans la terminologie de P. Lévy (1948).

La limite asymptotique de B' est C' = P/kT qu'on appellera capacité informationnelle sélective maximum de la classe de stratégies sélectives pouvant se rattacher au signal considéré, et des messages résultant de ces stratégies.

En intégrant par rapport à la durée thermique, l'on obtient le concept d'énergie E, et celui d'information sélective maximum:

C = E/kT ou entropie de Clausius.

(Remarque terminologique : il semble très fâcheux de suivre Shannon en appelant toujours entropie l'information sélective - et même des fonctions de gain qui ne sont pas de vraies informations - car ce sont en fait là des concepts différents. Mais il paraît opportun d'utiliser "entropie" avec divers qualificatifs appropriés, pour désignel diverses valeurs remarquables de H).

Dans le cas présent, l'entropie de Clausius n'est pas une nouvelle fonction d'état du système, car elle se déduit par multiplication de fonctions connues : énergie et information fishérienne spécifique de l'état du signal (§ 4.1.1). La forme de C explicite les analogies entre E/kT et 1/kT qui ont leur origine dans le fait que ce sont là deux informations, satisfaisant à des axiomes communs. Mais ce sont deux informations différentes et rien ne justifiait à priori l'identification à l'énergie du signal de leur rapport, qui était une énergie indéterminée avant l'introduction de la méthode actuelle de comparaison. La loi de Carnot donnera une valeur absolue à cette multiplication par E (§ 4.3.5) et permettra de définir la température (de fluctuation) comme paramètre d'équivalence de l'énergie et de l'information sélective, dans le cas où on peut maximer cette dernière parmi toutes les stratégies sélectives.

C = E/kT est une grandeur extensive, mesure d'une énergie extensive à l'aide d'unités donnant l'échelle des fluctuations et qui sont en fait des énergies intensives, liées à la notion de durée. C devait donc à priori avoir une signification profonde, induite par la durée thermique sur l'énergie (de même pour le quotient de l'action par le quantum d'action, grandeur induite par durée quantique sur l'action, qui se révèle être un entier).

4.3.3 - TEMPERATURE EQUIVALENTE ET ENTROPIE ABSOLUE

D'habitude, l'état du système ne permet pas à l'information d'atteindre le maximum C. Mais elle reste bornée par C, donc possède une limite supérieure B, que nous appellerons entropie de Boltzmann. (B s'identifiera à la "fonction de Massieu" de la Thermodynamique. Cf. Guggenheim (1949)).

B pourrait constituer un deuxième paramètre fondamental de l'état du système, s'ajoutant à la température. Mais on utilise plutôt d'autres paramètres, qui s'obtiennent en exprimant le défaut d'information de deux façons, l'une multiplicative, l'autre soustractive.

La première méthode conduira à la définition de la température équivalente, la deuxième à la définition de l'entropie absolue. Il résultera de la loi de Carnot (§ 4.3.5) que la température équivalente sera supérieure ou égale à la température de fluctuation, et que l'entropie absolue sera positive ou nulle.

Température équivalente. La première méthode consiste à écrire l'information potentielle sous la forme $B = E/kT\theta$. $kT\theta$ sera la température équivalente du signal. $1/\theta$ est le rendement, en temps nécessaire pour transmettre une information donnée, du signal donné par rapport au signal dégénéré d'entropie nulle.

Si par exemple l'état physique est tel que l'on ne peut pas dépasser la stratégie sélective de bande correspondant au bruit δ, $\theta = G^{-1}(P/\delta)$. S'il est impossible de donner au signal la distribution gaussienne, $\theta = (\delta/P) \log (A + P''/\delta)$ où P'' est la puissance entropique de Shannon, inférieure à P, qui aurait donné la même information avec le signal gaussien.

Entropie absolue. La deuxième méthode consiste à écrire l'information potentielle B sous la forme $B = E/kT - S = C-S$. S sera appelé entropie absolue du signal. Dans cette définition, on considère que la partie $Q = SkT$ de l'énergie est nécessairement redondante, parce que liée, et inaccessible aux opérations macroscopiques. Par définition kT_f est un facteur d'intégration pour Q.

L'information potentielle et l'entropie sont des expressions sans dimensions. C'était nécessaire, car ce sont des concepts relatifs à l'action de l'expérimentateur, qui peut toujours se ramener à une forme numérique : nombre de choix binaires. Il en résulte que l'on peut définir l'égalité des entropies de deux objets à températures différentes. Cette possibilité est très importante, car elle permettra d'introduire au § 4.4.4 un terme de comparaison des températures qui manquait à leur définition.

Exemple d'entropies. Si l'état physique est tel que la bande ne puisse pas dépasser la valeur conduisant au bruit δ, alors $S = E/kT\,(1-G(P/\delta))$. S croît d'abord lentement avec τ, puis rapidement au delà de la valeur pour laquelle $\delta = P$. Il serait intéressant de savoir s'il y a des cas, et lesquels, où le fait que l'entropie est non nulle peut effectivement être attribué à l'impossibilité de dépasser une certaine largeur de bande, impossibilité due par exemple à l'intervention de phénomènes nouveaux.

L'entropie n'est pas ainsi définie de façon entièrement constructive, car elle exige une limite, qui correspond à un codage fait par morceaux très longs, c'est-à-dire en pratique en codant tout le message en un seul morceau. On peut éviter d'avoir affaire à des limites (maxima) si l'on accepte d'avoir affaire aux ensembles conceptuels de signaux introduits par J.W. Gibbs. Alors l'entropie s'exprime comme différence entre deux valeurs du nombre d'opérations nécessaires pour connaître l'état du signal, c'est-à-dire du logarithme de la probabilité relative au choix de ce signal tout entier dans un certain ensemble conceptuel de signaux. La première valeur correspond au signal réel, la deuxième à un signal fictif de mêmes énergie et température.

"<u>Entropie de mélange</u>". On n'a plus affaire à des concepts fictifs lorsqu'on arrive à se ramener au cas discret. Ceci se produit lorsque des objets continus sont juxtaposés sans pénétration. Alors l'entropie diminue de l'information que présente leur configuration, qu'on appelle "entropie de mélange".

On peut considérer que l'entropie tout court est la limite de l'entropie de mélange lorsque les cellules diminuent autant qu'il est possible, c'est-à-dire jusqu'à un niveau déterminé par la température. La méthode que nous avons utilisée au § 4.3 fait précisément cela, et elle remplace les difficules discussions sur les entropies petit grain et gros grain.

On peut se demander s'il est possible de définir l'entropie sans limite, ni ensemble conceptuel, c'est-à-dire s'il existe un problème inverse menant à la limite de problèmes directs et fournissant un mode opératoire utilisable.

<u>Principe de Nernst</u>. L'un de ses énoncés affirme que l'entropie est partout la même au zéro absolu, donc que l'on peut la prendre comme nulle. Cet énoncé s'applique à la définition habituelle de S, qui comporte une constante additive indéterminée. S'il s'applique aussi à notre définition absolue, il signifierait que dans le cas où la Nature est sans défense du point de vue fishérien ($T = 0$) elle est aussi nécessairement sans défense du point de vue sélectif additionnel ($S = 0$) et on a un état purement mécanique.

Un autre énoncé de principe est que $T = 0$ est inaccessible. Si le 1er énoncé est admis, le 2ème résulte du principe qu'on ne peut pas transmettre de l'information sans support matériel. Sans le 1er énoncé, le 2ème n'aurait pas nécessairement résulté de ce principe additionnel, car on aurait pu concevoir que $S \to \infty$ et que l'état de $T = 0$ comporte une entropie de Boltzmann nulle. Alors il aurait pu continuer d'être accessible.

4.3.4 - PROPRIETES DES SIGNAUX CONCRETS

Les stratégies isostatiques ont un rendement en information égal à un, et l'on peut toujours restaurer l'état initial. Les transformations isostatiques sont donc réversibles au sens de P. W. Bridgman (1943). Cependant cet auteur n'a pas essayé de déduire de la restaurabilité les autres propriétés générales de la réversibilité, pour vérifier si ce sont effectivement celles qu'on lui attribue; ni de dégager les propriétés de réversibilité particulières. En fait, il y a aussi d'autres réversibili-

tés que la présente;donc nous allons dégager les propriétés particulières de celle-ci.

Stratégies isostatiques et fluctuations. Les signaux concrets à entropie nulle exigent $\delta \gg P$, c'est à dire que le signal doit être une fluctuation très probable à laquelle on fait correspondre un message.

Cette propriété appartient bien aux processus réversibles selon la thermodynamique habituelle. On peut aller jusqu'à dire que l'existence de fluctuations est nécessaire au concept même de réversibilité. On peut par exemple concevoir un processus de chauffage à volume constant, en isolant un gaz de la source de chaleur, au moment où il se trouve au faîte d'une fluctuation de température; on le branche ensuite sur une source à température T + dT avec laquelle il sera en équilibre. Si on veut détendre à T constant, on le détend d'une quantité telle que son état actuel soit une fluctuation assez probable de l'état futur.On pourrait sans changer la température, ajouter de l'énergie constituant une fluctuation probable, donc toujours de l'ordre de kT. De quelque manière que l'on opère, tous les échanges d'énergie seront de l'ordre des fluctuations d'énergie, si on veut que les fluctuations de température aient une probabilité acceptable. De plus, certains des processus énumérés sont des démons de Maxwell.

Petits signaux. Principe de Casimir (1945). Ce qui précède est indépendant de la valeur absolue de P. D'autre part, il n'y a aucun sens à dire qu'un signal est petit par rapport aux fluctuations, puisque celles-ci sont infinies.

Ceci est important pour une catégorie de problèmes considérés par Onsager, où, après avoir appliqué la thermodynamique macroscopique à certains phénomènes, l'on trouve à posteriori que ces phénomènes étaient en réalité microscopiques. Ce sera également le cas au Chap. 5. Casisir a fait remarquer qu'il fallait alors expressément postuler la validité des raisonnements faits sans précaution. Remarquons de notre côté que le principe de Casimir cesse même d'avoir un sens dans le cadre de la théorie présente, car la Thermodynamique fournit une échelle d'énergies - kT - mais aucune échelle de puissance.

(Une autre méthode de démontrer le principe est d'utiliser un modèle : ceci est conforme au principe de Guggenheim (1949): "Se reporter au substrat microscopique pour démontrer les formules devant lesquelles les deux principes macroscopiques sont impuissants".

Ce substrat microscopique comprend en particulier les phénomènes quantiques. La démonstration que le principe d'Onsager-Casimir est inclus dans celui de Guggenheim (quantique) a été faite à partir des équations de Schrödinger par Callen et Welton (1951). Réciproquement la plupart des conclusions que Guggenheim tire de "son principe" s'appliquent à des petites entropies, que la théorie macroscopique noie dans les fluctuations et ne permet donc pas de calculer.

Il serait intéressant de voir dans quelle mesure le principe de Guggenheim peut être remplacé par des principes phénoménologiques).

Vitesses des transmissions isostatiques.

On attribue généralement aux processus réversibles la propriété d'être infiniment lents. Si la largeur de bande peut devenir infinie, nos processus possèdent cette propriété, mais relativement à un rythme repère, lui-même infiniment rapide, et le bilan est une vitesse absolue finie.

En effet, on ne peut utiliser à fond l'information potentielle B que si τ est infiniment petit, donc le nombre d'opérations par seconde, qui est deux fois la "largeur de bande" W, est infiniment grand, et l'information par opération infiniment petite; donc le rythme est infiniment lent par rapport à W.

Mais rien ne limite en principe $C' = P/kT$, donc absolument, la vitesse peut être arbitraire. En d'autres termes, la condition de réversibilité isostatique en présence de bruit thermique seul ne limite pas la vitesse. Ce résultat est identique conceptuellement au fait démontré par Shannon que la capacité ne diminue pas quand le % d'erreur acceptable diminue. Par conséquent, Shannon aura démontré implicitement l'inexactitude dans ce cas d'un résultat généralement admis, à savoir que la réversibilité signifie lenteur infinie. Il était d'ailleurs somme toute choquant de baser toute la Thermodynamique sur des transformations ayant de telles propriétés. Tout au contraire, on se base ici sur les transformations qui font la meilleure utilisation de l'énergie, et, l'énergie étant à puissance constante, font le meilleur usage du temps.

4.3.5 - ETAT MACROSCOPIQUE - MINIMAX - LOI DE CARNOT

Nous venons de définir l'entropie qui, avec la température, donne les deux "paramètres d'état" annoncés au § 4.3.1. Ceci justifie le terme "isostatique" appliqué aux stratégies réversibles, qui satisfont au critère de minimax relativement à un certain ensemble de stratégies admissibles d'observation. Il se révèlera de plus en plus en détail que l'étude de ce minimax conduit à une structure identique à celle de la Thermodynamique. On peut en induire que la validité bien prouvée de cette science justifie aussi bien le critère de minimax que la fonction de risque égale à l'information apparente moins l'équivocation. Tout autre critère aurait conduit à d'autres définitions à la place de celles des entropies de Clausius, de Boltzmann et absolue, et à d'autres principes. Par suite, le critère de minimax est rendu plus vraisemblable encore qu'il n'était au § 4.2.2 où il ne résultait que de la forme fonctionnelle du bruit thermique. Quant au choix de la fonction de risque, qui résultait de sa liaison avec les décisions discrètes (comme toutes les décisions finales de la physique), il peut aussi maintenant être déduit de la Thermodynamique. Mais dans les deux cas ce rôle n'était pas évident.

Cependant, la théorie présente dépend de façon fondamentale de l'étendue de l'ensemble des stratégies admissibles. Or des phénomènes nouveaux, sub-thermiques, viennent toujours élargir cet ensemble de stratégies admissibles. Mais on a constaté expérimentalement que ceci n'enlève rien de sa valeur à la descrip-

tion par température et entropie. On fait par suite du caractère intrinsèque de la valeur du jeu limité un principe physique, la loi de Carnot, qui n'est en rien un principe de la Thermodynamique, mais au contraire un principe de couplage entre celle-ci et le reste de la physique, qui renforce encore le rôle privilégié du minimax et de notre fonction de risque à base d'information sélective.

Loi de Carnot. Il est impossible d'aller au-delà du maximum d'information relatif aux stratégies sélectives.

Si cette loi est vraie, tout résultat qui semblerait aller au-delà du minimax contiendrait une erreur grave, tout aussi grave dans le cas où il s'agirait de bruit inférieur au bruit thermique, que d'information sélective supérieure au maximum relatif aux stratégies sélectives. En particulier, tout démon de Maxwell à stratégie-repère minimax aura un rendement <1.

La durée sélective, identique à la durée thermique, était définie à partir du minimax, limité aux opérations thermiques. Postuler l'universalité de ce minimax implique donc que la "durée physique", définie par un minimax universel, est identique à la durée thermique.

4.3.6 - ETATS NON MINIMAX

Ils s'introduisent de deux manières : d'une part par des stratégies imparfaites de l'observateur; d'autre part, par tout nouveau phénomène conduisant à introduire des bruits autres que le bruit thermique (Cf. Lawson et Uhlenbeck, 1950, p. 79 ss), où la stratégie de la Nature dépend de celle de l'expérimentateur. Ce sont d'ailleurs ces stratégies qui permettent de parler de maximation de la part de la Nature.

Par exemple, le "bruit shot" donne une puissance disponible spécifique $\tau.\hat{\delta} = \frac{eI}{4R}$, où e est la charge de l'électron, I l'intensité du courant, A la limite de W infini, l'information potentielle sera $E\left(kT + \frac{eI}{2R}\right)^{-1}$. Il en résulte que l'effet de cette stratégie supplémentaire peut être annulé en prenant I très petit, donc la transmission très lente et le minimax continue alors d'être accessible. En d'autres termes, si l'on ajoute le bruit shot, la condition de réversibilité isostatique limite la vitesse de transmission.

(Si dans d'autres cas il se révèle impossible d'arriver au minimax, la réversibilité devient inconcevable. Si les phénomènes en question sont suffisamment importants et intrinsèques, il peut être intéressant de définir une réversibilité relative, et les températures équivalentes deviennent importantes.

4. 4 - TRANSFORMATIONS REVERSIBLES - CYCLE DE CARNOT

4.4.1 - CLASSIFICATION

Nous allons continuer de supposer que les stratégies sont les meilleures, étant donné l'état du système, supposé toujours définissable par température et entropie absolue. Mais, au lieu d'avoir à utiliser au mieux un état donné, il s'agira de passer

au mieux, à masse constante, d'un état $E_1T_1B_1S_1$ à un autre E_2T_2 B_2S_2. Le "principe de l'état initial et de l'état final" sera satisfait par définition. (Des difficultés n'apparaissent que s'il faut plus de deux variables d'état : alors il faut recourir à la théorie de Carathéodory et la réunir à la nôtre.)

Les transformations élémentaires à considérer d'abord se feront :

- soit à T constant : isotherme,

- soit à S constant : isotropique.

Ces transformations élémentaires resteront réversibles au sens de P.W. Bridgman, car elles conserverons le rendement optimum en information des énergies utilisées. Mais elle différeront profondément des transformations isostatiques. En effet, bien que ces dernières soient tout à fait irréelles, elles sont construites comme limites, et on en étudie à posteriori les propriétés, et en particulier la vitesse. Au contraire, les transformations présentes, tout en continuant d'exiger la limite précédente dans la définition des états initial et final et des états intermédiaires, ne sont elles-mêmes que définies par des propriétés qu'elles doivent satisfaire. Elles ne peuvent être construites que si l'on connaît les propriétés physiques du signal. En particulier, leur vitesse exigerait de connaître les propriétés précises du signal.

Les échanges porteront sur de l'énergie libre, portant l'information E/kT, et sur de l'énergie liée, ne portant aucune information, et ajoutant à l'entropie du signal.

L'énergie libre, donc son information, peuvent être utilisées en travail, par exemple par des mécanismes du type de Szilard (1929). L'énergie liée ne porte pas d'information et ne peut donner de travail. Sous la forme où elle se présente dans les échanges, nous l'appellerons par définition chaleur. Donc tout comme dans l'axiomatique habituelle de la Thermodynamique, la chaleur est définie de façon indirecte, comme différence entre la fonction d'état énergie totale et le travail.

Il est évident que l'énergie n'est pas libre et liée per se, mais seulement par rapport à un observateur. Les transformations élémentaires ont pour but la construction d'un cycle de Carnot qui permette de faire passer l'énergie d'un état à l'autre.

Pour cela, il faut considérer à la fois des sources d'énergie libre et d'énergie liée à des températures différentes : la chaleur à haute température est plus "noble" que celle à basse température, parce que le fait de "lier" une certaine énergie la dégrade moins, et diminue moins l'information potentielle lorsque cette énergie est à haute température.

4.4.2 - TRANSFORMATIONS ISENTROPIQUES

Voyons comment l'équilibre entre les deux parties de l'énergie peut changer avec l'état physique : de l'énergie liée et inutilisable dans un état devient transformable en travail dans un autre état. Le paramètre essentiel de l'état est kT_f. Supposons qu'il décroisse de T^+ à T^-. Des signaux dont les amplitudes paraissaient égales, qui par suite ne pouvaient transmettre au-

cune information et étaient chaleur, deviennent susceptibles d'être distingués et on peut leur confier de l'information. D'autres signaux, qui transmettaient de l'information, peuvent en transmettre plus.

(En somme, la dimension minimum kT des cellules macroscopiquement inanalysables (§ 4.1.7) a diminué, et l'intérieur des cellules kT^+ est devenu analysable à la température T^-).

Par suite, si le passage de T^+ à T^- a été fait de façon adiabatique, c'est-à-dire sans échanges d'énergie liée (chaleur), une certaine énergie libre a été rendue disponible, et si l'information à transmettre n'a pas non plus changé, une partie de cette énergie peut être extraite et emmagasinée sous forme de travail (Cf. § 5.2.3). Réciproquement, pour passer de T^- à T^+, il faut ajouter de l'énergie, fournie par exemple par du travail.

Si à chaque fois que la température change le recodage est fait de telle façon que tout signal continue toujours à transmettre le maximum d'information, nous dirons que la transformation est isentropique.

4.4.3 - TRANSFORMATIONS ISOTHERMES

Par définition même de l'état macroscopique du § 4.3.5., une transformation isotherme et non isostatique exige à la fois des échanges d'énergie libre et d'énergie liée.

4.4.4 - COMBINAISON DE STRATEGIES - CYCLE DE CARNOT

Soit un cycle formé de deux transformations isothermes et deux isentropiques entre T^+ et T^-, S^+ et S^-.

Tout comme l'information libre, l'entropie échangée doit être nulle au bout du cycle; il faut donc que $Q^+/T^+ = Q^-/T^-$, ce qui veut dire qu'une énergie libre $E = Q^+ - Q^- = Q^+\left(1 - \frac{T^-}{T^+}\right)$ a été récupérée en travail.

Le cycle qui précède est bien entendu identique à celui de Carnot. Il permet de conclure que la température T_f de fluctuation introduite au § 4.1 est identique à la température T_k de Kelvin, c'est-à-dire lui est proportionnelle. Cette dernière résultait également de la transitivité d'un équilibre, l'équilibre thermique (si A = B et B = C; alors A = C; - Cf. Landé 1926, p. 284); mais elle pouvait ne pas être identique à la température de fluctuation. La définition de T_k exigeait d'ailleurs une dimension nouvelle, exprimée par k, qui n'était déterminée que lorsque l'on constatait l'identité de propriétés entre deux définitions différentes de l'entropie, dont l'une est numérique. Ce paramètre k revenait au choix d'un cycle de Carnot unitaire, marginal, où $\Delta S = 1$.

4. 5 - TRANSFORMATIONS IRREVERSIBLES QUELQUES DÉMONS DE MAXWELL

4.5.1 - IRREVERSIBILITE

On appellera irréversibles toutes les stratégies qui n'atteignent pas les optima correspondant aux stratégies isostati-

ques, isothermes ou isentropiques, ou des combinaisons de celles-ci.

Alors, une partie du signal perd sa valeur informative : un codage irréversible signifie donc qu'une partie de l'énergie initialement disponible a cessé d'être susceptible de former un signal, devient liée.

A puissance de signal donnée, l'irréversibilité signifie que la durée a été mal utilisée. Or l'essence même de la durée est d'être orientée : on ne peut pas revenir en arrière pour améliorer l'utilisation du passé.

Toute mauvaise utilisation de la durée traduit une désadaptation entre stratégie et son objet, ou une mauvaise réaction face à une opposition donnée de la Nature. Nous avons vu trois types de désadaptation :

(1) - Perte de toute espèce d'information si l'action est retardée -

(2) - Impossibilité d'atteindre le maximum d'information fishérienne : mauvaise utilisation de la durée thermique -

(3) - Impossibilité d'atteindre le maximum d'information sélective : mauvaise utilisation de la durée sélective.

En fait, toute action comporte chacune de ces trois pertes (elles deviennent nécessaires dès que l'on tient compte du coût des raisonnements qui conduisent à l'adaptation !). L'existence nécessaire de ces pertes peut être posée en principe, qui contiendrait une partie du 2ème principe de la Thermodynamique. Ce sont des aspects fondamentaux et irréductibles de la durée, et la traduction la plus commode de l'orientation du temps.

A ce principe se rattachent les hypothèses d'"irréversibilité des processus de mesure", ou d'"irréversibilité de toute évolution", la dégénérescence, l'interprétation probabiliste de l'entropie, etc... Tous ces énoncés sont également incapables de fournir une structure de la durée. Nous renoncerons donc à les mettre au premier plan de la définition de la durée intrinsèque de la Thermodynamique. Celle-ci sera dès lors définie :

- ou bien comme une durée sélective, quotient de l'énergie par le maximum d'information sélective transmissible, si l'entropie est nulle -
- ou bien comme identique à la durée thermique -
- ou bien comme durée quantique, à travers une théorie cinétique.

On voit donc que si la Thermodynamique classique semble ignorer la notion de temps, c'est de façon superficielle, car elle introduit la notion de transformation réversible et celle-ci n'est physiquement consistante que si l'on admet l'existence de fluctuations, si petites soient-elles, donc du temps.

L'utilisation d'une stratégie irréversible dégrade le signal par rapport à l'observateur considéré. C'est la conséquence du fait que les signaux utilisés restent nécessairement sous la

forme sous laquelle ils ont été utilisés, tandis que le stock aurait pu être utilisé différemment. (Cependant, tout principe de conservation ou d'évolution n'ayant de sens que par rapport à un cycle fermé et un observateur, on ne peut dire qu'il y a dégradation universelle, sans par cela même postuler l'existence d'un observateur de référence universel.

4.5.2 - SIGNAUX ABSTRAITS

Nous dirons qu'un signal provenant de stratégies irréversibles est abstrait : la correspondance entre signal et message constitue un filtrage plus ou moins poussé pour en extraire l'information, à laquelle certaines mesures n'apportent rien.

Le recodage optimum (Chap. 5) est celui qui fait passer tout message à l'état de message concret, quelle qu'ait pu être sa nature préalable : abstrait ou concret, du point de vue d'un autre système de codage, c'est-à-dire d'un autre mode d'étude des propriétés physiques.

La distinction fonctionnelle abstrait-concret peut remplacer la distinction formelle faite par Shannon entre signaux discrets (arithmétiques) et continus (analytiques).

Un signal discret n'est pas nécessairement abstrait. Contre exemple : le véhicule du signal est quantifié physiquement. Un signal continu n'est pas nécessairement concret. Contre-exemple : signaux envoyés sur une voie téléphonique : l'organe des sens est un appareil de mesure qui filtre progressivement une partie énorme du signal continu, et rend le signal discret (et concret ?).

De toute façon, la propriété d'un signal d'être continu ne peut être sa propriété essentielle. Seules des raisons pratiques, et non des raisons théoriques, rendent indispensables l'étude des signaux continus, qui au Chap. 5 sera faite du point de vue pratique des décodeurs quantificateurs.

Réciproquement, un signal abstrait n'est pas nécessairement discret, ni un signal concret continu.

Le code qui est le plus économique lorsque l'on connaît la clef conduit à un signal concret qui est indéchiffrable lorsque l'on ne connaît pas la clef. Employer un tel code c'est donc beaucoup demander au décodeur, que la moindre désadaptation rendrait alors inutilisable; il perdrait trop en souplesse en contrepartie de sa perfection. En pratique donc, le signal devra aussi contenir de l'information sur la clef du décodage, ce qui se traduira par de la "redondance" interne, et de l'irréversibilité. L'exemple extrême de redondance tout-à-fait localisée est celui où la définition d'un mot ou d'un signe rare accompagne ce mot ou le signe.

4.5.3 - RENDEMENT. REDONDANCE. AUGMENTATION D'ENTROPIE

Les définitions du § 4.3.4 peuvent être également écrites lorsque la responsabilité de la limitation de l'information ne provient pas de l'objet mais du fait que l'ensemble de stratégies est restreint (ou même réduit à une seule stratégie imposée).

On définit ainsi le rendement de la stratégie, en mesurant l'irréversibilité par le rapport du temps utilisé au temps écoulé, ou de l'information reçue à l'information potentielle. L'ordre créé représente la partie réversible du processus. Un moins le rendement est la redondance.

Ce rendement est calculé par rapport à la meilleure stratégie compatible avec l'état du système. Il faut donc le multiplier par celui du système lui-même, si on veut prendre comme repère l'énergie du système et son information maximum au lieu de l'information potentielle de son énergie.

Par exemple, la limitation de l'information sélective peut être due au fait que la mesure du signal avait été faite avec erreur systématique b(x). Alors l'égalité de l'information de Fisher et de l'inverse du carré de la variance est remplacée par l'inégalité de Fréchet-Darmois (retrouvée plus tard par Cramer et Rao) :

$$\frac{1}{\sigma^2} \leqslant n \frac{\text{inf. de Fisher}}{(1 + B'(x)^2)}$$

d'où l'on peut déduite une limite inférieure de θ.

La température équivalente a été définie par exemple par Lawson et Uhlenbeck (1950, p. 88). Nous avons aussi défini une "température informationnelle", dont la théorie sera donnée au Chap. 7 : la limitation sera due au fait que certaines opérations doivent être réservées à des fins non-informatives, les autres restant parfaites, et les proportions relatives des deux restant indépendantes de T. Alors, si la température T est uniforme, elle est en facteur, et on peut la supprimer. La température relative hérite alors du rôle de facteur d'équivalence entre information potentielle et information que jouait T. Mais la température informationnelle n'est pas elle-même un rendement.

Il est intéressant de faire remarquer ici qu'en Thermodynamique Relativiste (Cf. Tolman, 1933) on utilise le postulat que l'introduction de la considération du mouvement n'étend pas l'ensemble où B' est maximé, et que par suite C' reste le même qu'en l'absence de mouvement. Comme $Q = Q_0 \, (1 - v^2/c^2)^{1/2}$ il faut que $T = T_0 \, (1 - v^2/c^2)^{1/2}$; soit $\theta = (1 - v^2/c^2)$.

Par ailleurs, l'irréversibilité peut être mesurée par l'augmentation d'entropie qui accompagne l'utilisation de la stratégie considérée.

4.5.4 - DEMONS DE MAXWELL

Les démons de Maxwell dont la stratégie repère est réversible sont de loin les plus importants et les plus proches de l'esprit du démon original. Ils présentent une particularité du fait même de leur définition. En effet, la classe réversible a été définie par des opérations de maximation, dans un domaine de plus en plus étendu, et par suite aucune question ne se pose quant à l'impossibilité d'un rendement supérieur à 1 pour un démon dont le deuxième terme appartient à ce domaine de maximation. Ceci permet de se placer dans la première de nos perspectives générales d'étude.

4.5.5 - DEMONS DE MAXWELL A REPERE REVERSIBLE ET STRATEGIE UTILISEE OPTIMALE

Démon de Szilard (1929). Les principales considérations sur l'impossibilité d'un rendement > 1 pour un démon de Maxwell sont dues à Szilard. Certains de ses modèles sont réalisables, bien que très théoriques, mais il a également introduit un démon limite, de rendement égal à 1, qu'aucun démon réel ne saurait surpasser (et que nous aurions pu étudier dès le § 4.3).

Nous reprendrons ce démon au chapitre 5, mais on peut tout de suite remarquer qu'en fait, ce n'est pas un vrai démon à deux termes différents. Les deux stratégies sont toutes deux réversibles, et la différence apparente provient de ce que la stratégie réversible repère n'est pas explicitée, mais son résultat estimé par le bilan entropique. D'après notre définition de la réversibilité et de l'entropie,et le caractère optimal de la stratégie utilisée,le bilan de celle-ci est identique au bilan entropique. Finalement, en somme Szilard n'a fait que comparer deux mesures d'un même changement : l'une abstraite (le repère), l'autre relative à un procédé qui serait concret, s'il n'était limite. Le résultat est un démon donc par définition même le rendement peut atteindre l'unité, mais ne peut la dépasser.

Démon de Smoluchowski -1912). La stratégie repère est toujours réversible. La stratégie utilisée est thermique, et d'ailleurs très mal explicitée. Mais de toute façon elle conduit au moins à augmentation d'entropie.

L'étude des deux démons précédents a contribué de façon essentielle à l'ensemble de ce travail.

Le démon de Maxwell de L. Brillouin (1951).

Cet instrument ne devrait pas, strictement parlant, figurer ici, car il sort de la Thermodynamique non quantique. En fait, il introduit une explication quantique de l'augmentation d'entropie consécutive à la stratégie utilisée, donc en fin de compte une explication quantique de la durée thermique. Ce démon fait donc partie des fondements quantiques de la Thermodynamique.

4.5.6 - LES DEMONS REVERSIBLES REALISABLES

Leur rendement est strictement inférieur à 1, et ils comportent nécessairement un certain pourcentage d'erreur qui rend le bilan compliqué, car l'équivocation doit faire partie de la chaleur non compensée.

Démon sélectif-thermique quantificateur. (Cf. Lawson-Uhlenbeck p. 168).

Considérons par exemple un démon du type de la modulation à impulsions codées (PCM) (qui se rapproche du modèle de Maxwell réduit au problème de reconnaître si l'amplitude d'un signal dépasse ou non une valeur donnée).

Soit donc un signal à deux valeurs $P = h\delta$ et $-P$, auquel se superpose un "bruit" normal de Laplace-Gauss d'écart δ. Notre stratégie sera d'identifier à $+P$ tout signal reçu positif et à $-P$ tout signal négatif. La probabilité d'erreur est $p = p(h)$. L'information reçue est en apparence $1/\tau = 2\delta/kT = 2P/hkT = 2S/h$. Elle pourrait devenir supérieure à l'information poten-

tielle $S \log_2 e$ si $h < 2 \log_e 2 = 1{,}3862$, c'est-à-dire si la probabilité d'erreur admissible était $p \geqslant 0{,}084$.

Ce démon de Maxwell est donc capable d'échanger la sécurité contre un excédent d'information sélective. Calculons l'"équivocation des résultats" $E = - p \log p + (1 - p)$ lot $(1 - p)$ qui permet de comparer ces deux notions. L'inégalité $h/2 \log_e 2 > 1 + p \log p + (1 - p) \log (1 - p)$ montre que l'échange qu'effectue ce démon a au plus un rendement égal à 1. Le gain d'information sélective qu'il semble obtenir au-delà de l'information potentielle est donc illusoire.

Si l'information à transmettre par les signaux usiels peut être considérée comme sélective, l'unique problème théorique de la technique de la modulation est le problème direct de l'étude du meilleur deuxième terme d'un démon à premier terme réversible.

Tous ces démons qui diminueront l'information potentielle et introduiront des erreurs,rendront cependant cette information accessible en la faisant passer à un niveau d'analyse beaucoup plus gros.

En général, la croissance de l'équivocation avec la probabilité de perte est trop rapide : nous verrons au § 6.1 un exemple (procédé Zator) où l'erreur est introduite à dessein. Le bilan ci-dessus n'est donc pas entièrement désavantageux.

Un problème technique extrêmement significatif consisterait, pour chaque groupe de techniques déterminé par l'importance relative du bruit, à construire un premier terme de référence non réversible, c'est-à-dire à considérer des rendements non intrinsèques, mais pratiques, permettant de comparer les procédés de modulation à un idéal moins irréel que la stratégie réversible. L'introduction de cette nouvelle stratégie équivaut à redéfinir la capacité informationnelle des lignes selon Shannon. (Ceci correspond d'ailleurs au fait que dans certains cas l'information sélective ne correspond plus à l'information intuitive du message).

Théorèmes H. Notre forme de la loi de Carnot ne peut pas être considérée comme démontrée par des théorèmes du type "H" : ceux-ci sont du type de l'impossibilité des démons de Maxwell, et s'appliquent à des mécanismes particuliers. En effet, le plus souvent, on cherche explicitement des expressions qui augmentent dans des processus déterminés. Le procédé habituel est d'introduire des cellules, des prises de moyenne, etc... En particulier, R.G. Tolman (1938, p. 172) a donné un théorème H généralisé basé sur une distinction entre une entropie grain fin qui reste constante et une entropie gros grain qui croît nécessairement. Dans ce cas l'introduction des cellules est identique à la stratégie thermique quantificatrice et la démonstration de Tolman se ramènera à notre démonstration de l'impossibilité du démon à 2° terme thermique quantificateur où l'on aurait négligé tout aspect équivocation, par exemple en le supposant constant. Donc le théotème de Tolman est involontairement extrêmement particulier.

CHAPITRE 5

THERMODYNAMIQUE DU DÉCODEUR

5.1 - ADAPTATION DU MESSAGE AU DÉCODEUR

5.1.1 - DEFINITION

La théorie inverse que nous avons jusqu'ici développée se caractérisait par le fait que nous nous occupions uniquement des limites à la stratégie d'encodage, sans nous soucier du message, que nous supposions toujours le plus favorable qui soit.

Nous allons donc maintenant aborder le problème de la recherche des propriétés du message discret adapté à la transmission sur un support continu, et au décodage au moyen d'une stratégie réversible. Cette étude ne sera pas en réalité autre chose qu'une adaptation au cadre présent d'une théorie fondamentale de L. Szilard (1929). Son principal intérêt sera de créer un pont entre le continu, représenté au chapitre précédent, et le discret, étudié plus en détail au chapitre suivant. (Sur ce passage, et la signification de l'élimination de la Nature, voir § 0.3.4 fin).

La méthode de ce chapitre et du suivant est "microscopique". Ce terme signifie que le message est considéré dans ses éléments : c'est une suite de tirages au sort entre R éléments M_n $(1 \leqslant n \leqslant R)$, dont les probabilités p_n caractérisent le message. Ce point de vue statistique dans le problème de la transmission de l'information est surtout dû à Kolmogoroff (1941) et Wiener (1948).

A cause du rôle particulier de la langue et de l'insuffisance des termes physiques et physiologiques, il nous a paru commode d'utiliser le vocabulaire de la linguistique, là où de nouveaux termes devaient être introduits. C'est ainsi que nous appellerons toujours les éléments M du message, les "mots du message". Le sens très étendu de ce terme ne doit pas prêter à confusion. Plus généralement, pour faciliter l'exposé, nous l'avons axé sur le problème linguistique, même en ce qui concerne la relation formelle d'adaptation. Il va cependant sans dire qu'on n'a introduit aucune hypothèse spécifiquement linguistique et que, par suite, ce qui suit sera une théorie générale du problème d'adaptation statistique, applicable dans tous les cas où,

par suite du rodage entre signal et message, des signaux "naturels" sont produits de façon à ce que leur décodage par les récepteurs naturels puisse se faire avec le meilleur rendement en information.

5.1.2 - DECOMPOSITION

Considérons un processus de transmission sans réaction (feedback) d'un élément sur les précédents et sur les suivants. Nous le dirons "ouvert". S'il y avait réaction, nous le dirions "à retour". Supposons qu'il puisse être parcouru dans les deux sens du point de vue informationnel. Nous le dirons "réversible du point de vue informationnel". Tous les éléments du processus seront dits "passifs", par extension du sens électronique du mot : ils n'apportent d'information dans aucune direction.

Nous étudierons ici deux catégories de blocs du point de vue fonctionnel et opérationnel :

- (D) différenciateurs, ou identificateurs, ou multiplicateurs, ou opérateurs d'expérience.
- (D^{-1}) intégrateurs, ou opérateurs d'action.

La suite d'un (D^{-1}) et d'un (D), soit (D^{-1}.D) représente un morceau de ligne à codage unique.

La suite d'un (D) et d'un (D^{-1}), soit (D.D^{-1}) fait passer d'un véhicule de message à un autre véhicule. Son bilan entropique ΔS doit être positif.

Par définition, nous dirons qu'il y aura adaptation fonctionnelle (entre expérience et action) si ce ΔS est minimum, étant donné les principes du bilan et toutes autres contraintes physiques qui ont pu être introduites.

5.2 - ÉLÉMENTS FONDAMENTAUX (D) ET (D^{-1}) ET BILAN

5.2.1 - DIFFERENCIATEUR

Nous appellerons (D) le passage d'une représentation du message M comme concept synthétique, à une représentation comme suite de mots M_n : par exemple, le découpage d'un code en tranches et leur identification : la possession d'un (D) constitue donc la mémoire.

(D) est l'analogue du "processus (1)" de la Quantique selon von Neumann (1932) ou du processus de séparation des éléments d'un mélange de Szilard (1929). Il s'agit donc, d'une part, d'un changement de point de vue, l'attention passant du message entier à un fragment de son véhicule. Mais ce changement ne peut se faire sans coût, car il exige une opération physique effective : l'identification de l'état M_n sans destruction. C'est une multiplication du message et elle s'accompagne de l'augmentation S_n de l'entropie de l'ensemble du système.

Nous ne discuterons pas les origines possibles de cette entropie, qui devrait de préférence provenir d'une dégradation apparente d'une source d'énergie libre à la disposition de (D). Nous nous contenterons de chercher une limite inférieure à S_n, qui à priori aurait pu être nulle.

Posons $S_n = k \log_e q_n + S_o$ où $\sum_n q_n = 1$

donc $S_o = - \log \sum_n \exp (-S_n/k)$

Les q_n seront appelés les pseudo-probabilités caractéristiques ou propres de (D). Une telle définition est inverse de celle de Boltzmann, qui part des probabilités pour arriver à l'entropie (Cf. § 4.3.3).

Le bilan de (D) est en somme :

$$S = S_1 = - k \sum_n p_n \log_e q_n + S_o$$

(Les q_n jouent le rôle des valeurs propres des opérateurs hermitiens de la Quantique. Ces valeurs propres, ainsi que les "vecteurs propres" M_n, sont des caractéristiques de (D), qu'ils décrivent entièrement de notre point de vue. Par analogie avec la Quantique, l'on pourrait dire, pour définir l'"état pur" M_n, que c'est l'état tel que l'application de (D) se réduit à conserver ce mot inchangé; mais ici cela n'ajouterait rien d'intéressant, car si (D) ne change pas l'"état" abstrait, il change bel et bien l'"état" physique du système, en donnant au véhicule du signal la possibilité d'agir sur la suite du processus de communication. Par suite, l'analogie entre Quantique et Théorie stratégique du (D) ne peut être complète dans les interprétations).

5.2.2 - INTEGRATEUR

(D^{-1}) sera le processus inverse de (D), passant de M_n à M. La possession de (D^{-1}) constitue l'intelligence. (D'autre part, le processus d'émission peut être interprété comme (D^{-1})).Donc, contrairement au processus physique (D), (D^{-1}) est uniquement un processus conceptuel de changement de point de vue. Ce ne signifie point qu'il soit sans signification physique, car la notion d'entropie dépend précisément du point de vue où l'on se place. Tout principe de variation de l'entropie ne peut être valable qu'à un point de vue donné. Dans le changement de point de vue intervient l'"entropie de mélange" (Guggenheim (1949) et le bilan de (D^{-1}) est :

$$S = S_2 = - k \sum_n p_n \log_e p_n$$

5.2.3 - CRITERE DE (D)(D^{-1})

Dans le jeu Emetteur contre Nature du Chapitre 4, la fonction de gain (= gain apparent - fonction de risque), ainsi que le critère, étaient déterminés par le désir de retrouver des résultats connus expérimentalement. Il en sera de même au Chapitre 6.

Mais dans le jeu présent, nous avons pu poser à priori une définition de l'adaptation. Elle consiste à s'adapter à un message donné, ou tout au moins attendu, et non pas au message le plus défavorable. C'est donc une Solution Bayes au sens de A. Wald (1950). La fonction de gain est également posée à priori : c'est $- \Delta S = S_2 - S_1$ différence entre une fonction de gain et une fonction de coût.

Pour que $-\Delta S$ soit maximum, il faut $\boxed{p_n = q_n}$ donc qu'il y ait égalité entre probabilités de (M) et pseudo-probabilités de (D). Ceci justifie ce dernier terme.

Par définition de q_n, s'il y a adaptation, $(D.D^{-1})$ se traduit par l'augmentation d'entropie S_0. Celle-ci devant être positive, nous obtenons la <u>relation de Szilard</u> (1929) :

$$\Sigma \exp(-S_n/k) \leqslant 1,$$

qui donne une limite inférieure d'<u>ensemble</u> aux S_n : elle joue un rôle de relation décompensation par laquelle un S_n isolé pourrait être très petit, à condition que les autres soient très grands. Elle montre que la loi de Carnot donne une limite inférieure au coût de reconnaissance, dont S_n fait obligatoirement partie. Au Chapitre suivant, nous laisserons de côté les détails du Chapitre présent, dont nous ne conserverons que ce résultat : <u>impossibilité de transmettre sans coût</u>.

Si $S_0 = 0$ il y a réversibilité thermodynamique aussi bien que réversibilité informationnelle : la mesure concrète du message par les S_n s'identifie alors à sa mesure abstraite par les p_n.

On peut supprimer de tels $(D.D^{-1})$ ou $(D^{-1}.D)$ et par suite accoupler de tels (D) et (D^{-1}), contigus ou non, en supposant les autres couplages parfaits (d'où l'avantage de ces concepts pour l'analyse détaillée de la transmission).

Si nous représentons un flot d'information d'amont en aval, vers des états d'entropie croissante, des couples $(D.D^{-1})$ de $S_0 = 0$ représentent des paliers horizontaux, d'entropie constante. Réciproquement, on peut intercaler de tels paliers horizontaux au milieu de tout processus (à condition de ne pas poursuivre l'analyse jusqu'aux effets quantiques). On peut même remarquer qu'un processus isentropique (§ 4.4.2) peut être considéré comme limite continue d'une suite alternée de (D) et (D^{-1}) adaptés, où chaque tronçon $(D^{-1}.D)$ serait à une température différente du précédent. Alors, en plus du bilan entropique, il faut considérer un bilan énergétique.

Remarquons aussi qu'on peut devoir tenir compte, dans le bilan entropique, des valeurs inégales que représentent, pour qui régit la transmission, les entropies dépensées en divers points du circuit. En effet, si on ne peut point espérer avoir des mécanismes tout entiers d'un bon rendement, on peut avoir à se résigner à ce que les parties les plus exposées, les plus robustes et les moins fines aient un rendement médiocre. Supposons que l'on veuille rendre minimum les pertes, pondérées des valeurs relatives de ces entropies :

Si p_n est donné, on retrouve $q_n = p_n$, mais si q_n est donné, il faut, pour que $f(S_1) - g(S_2)$ soit minimum, que l'on ait :

$$-f'(S_1)(\Sigma \Delta p_i \log p_i) + g'(S_2)(\Sigma \Delta p_i \log q_i) - K(\Sigma \Delta p_i) = 0$$

d'où :

$$p_n = Q\,(q_n)^B \quad \text{avec} \quad B\,\frac{f'(S_1)}{g'(S_2)} = V(B)$$

Cette équation détermine $B \geqslant 0$. Elle pourrait avoir plusieurs solutions ou aucune; dans le cas linéaire $f' = B_1$; $g' = B_2$; $B = B_1/B_2$.

5.3 - RÉALISIBILITÉ DE (D) RÉDUCTION A UN PROBLÈME SÉQUENTIEL

(D^{-1}) est toujours réalisable. Par contre, (D) doit être physiquement envisagé comme étude expérimentale des propriétés du véhicule du signal, et il se peut qu'il soit impossible de construire un appareil qui ait une caractéristique (q_n) arbitrairement donnée.

Par ailleurs, il existe des cas où il est impossible d'étudier à fond cette réalisabilité sans passer à des détails de structure trop spécifiques. En particulier, si les opérations de (D) sont indécomposables en opérations plus simples, l'étude ne peut être poussée plus loin et la seule relation entre les S_n est la relation de Szilard. Or un mode opératoire expérimental ne peut avoir de sens physique que si l'identification complète d'un signal exige un nombre fini f d'opérations, donc chacune a un nombre fini q de résultats possible. Par suite, l'indécomposabilité de (D) entraîne que le nombre R de mots est fini : d'ailleurs en pratique il doit être petit ou modéré.

Si par contre (D) est décomposable, le problème spécifique se trouve reporté aux opérations élémentaires, tandis que la composition de celles-ci peut être étudiée indépendamment des détails de structure. L'analyse qu'effectue cet zppareil se présentera alors de la façon suivante qui fera du mode opératoire expérimental de (D) une "fonction de décision séquentielle" au sens de Wald (1947) : avant que l'expérimentation commence, l'incertitude sur le numéro du mot est totale. Après chaque expérience, la fonction de décision permet de prendre l'une des deux décisions suivantes : continuer à expérimenter (décision g, où $1 \leqslant g \leqslant q$) ou arrêter en identifiant le mot étudié (décision 0). La suite des expériences identificatrices constitue un code pour le système à identifier. Si le coût de l'expérience E_g est indépendant de son résultat g, la fonction de coût au sens de Wald sera dite <u>simple</u>. C'est le cas le plus étudié par Wald. (J. Ville et M.P. Schutzenberger (1951) l'ont indépendamment abordé dans le même contexte). Si le coût dépend du résultat de l'expérience, la fonction du coût sera par extension dite <u>semi-simple</u>.

Le meilleur mode opératoire est celui qui donne le coût le plus économique (par exemple, dans le cas simple, celui dont le nombre d'expériences moyen est le plus petit).

Ceci exige d'abord des signaux indépendants, donc degrés de liberté indépendants et stratégie séquentielle invariable depuis le début jusqu'à la fin de l'identification, indépendamment des résultats partiels.

Par ailleurs, on peut supposer illimité le nombre des opérations pour une seule identification. Alors, dans la mesure où les délais de décodage sont illimités, le nombre de mots le sera lui aussi.

Cependant, la fonction du coût étant la somme des coûts élémentaires, les S_n ne pourront plus être indépendants les uns des autres. Mais dans l'étude de cette dépendance, les aspects physiques, fonctionnel et thermodynamique n'interviennent plus. On a affaire à un problème abstrait de codage séquentiel, c'est-à-dire, de théorie probabiliste des processus aléatoires ponctuels. Elle fera l'objet du chapitre suivant, où seul le concept de coût subsistera du chapitre présent.

L'on voit dès lors que la théorie de l'information sélective est ce que l'on obtient lorsque l'on étudie les stratégies séquentielles optima de façon formelle, indépendamment de leur origine stratégique. Celle-ci ne subsiste que par le caractère physique de la durée physiologique de (D) qui résulte de ce que, du fait de la permanence de (D), les opérations élémentaires faites à des instants différents ne doivent pas différer entre elles. Si ces opérations sont tnermiques, la durée physiologique sera identique à la durée thermique, donc du fait du principe de Carnot, à la durée physique.

Le rôle de la Stratégique dans la physique est donc finalement tout différent de celui de la théorie de l'information : elle est plus abstraite que cette dernière discipline, qui peut être considérée comme une de ses réalisations physiques concrètes.

5. 4 - ÉTUDE DE L'ADAPTATION

5.4.1 - ADAPTATION ET ERGODICITE

L'adaptation se traduit donc par les égalités :

$$S_n = - k \log_e p_n$$

Rapprochons ces relations des relations de Botlzmann où S_n est défini à partir de la probabilité W_n de la configuration du véhicule physique correspondant au signal. On aura :

$$k \log_e W_n = k \log_e p_n \text{ , soit } W_n = p_n$$

Ecrite de cette façon, la relation d'adaptation Bayes prend un caractère ergodique : une fonction d'un espace conceptuel à temps constant est égale à une fonction d'un espace réel à temps variable. Cependant, les difficultés auxquelles mène l'espace conceptuel de Boltzmann sont très connues, et ont été abordées au § 4.7. Khintchin (1949) déclare même ne pas voir clairement ce que signifie "la probabilité W_n". Nous nous demandons par suite s'il n'est pas souhaitable d'abandonner cette construction d'espace conceptuel à priori, et de considérer l'ergodicité ci-dessus comme une condition physique à satisfaire à posteriori.

5.4.2 - PROBABILITES SUBJECTIVES

La définition des p_n comme limites de fréquence peut alors être remplacée par l'hypothèse d'adaptation ajoutée à la description du décodeur. Il semble bien que c'est là exactement ce que l'on fait en réalité avec toutes les définitions de la probabilité faisant intervenir des éléments subjectifs : tels que "aucune raison de croire que p_1 et p_2 sont différents". L'opération (D) plus l'hypothèse d'adaptation sont donc une représentation de ce que l'on veut dire par ces définitions subjectives. Cette représentation remplace par des actions (D), les jugements qui sont d'habitude à la base de telle définitions.

5.4.3 - PROBABILITE D'UN EVENEMENT UNIQUE

Supposons que le nombre total des mots soit petit. Il n'existe alors aucun moyen de mesurer les probabilités comme limite

de fréquences, donc on n'a aucune possibilité de vérifier l'adaptation,et par suite aucune notion d'entropie de mélange récupérable. La seule manière naturelle que l'on ait alors d'estimer la quantité d'information du mot, est de supposer qu'il y a adaptation et d'attribuer le message à un tirage au sort dans un ensemble adapté au mécanisme décodeur, c'est-à-dire ayant des probabilités pour les mots égales aux pseudo-probabilités propres à (D). De plus, la quantité d'information du message est estimée par la "capacité" de (D) :

$$K = - \sum_n q_n \log q_n$$

Il existe un cas où ceci est fait nécessairement. C'est celui où le message est unique par nature, par exemple dans certains jeux, dans les élections, courses, etc... Les pseudo-probabilités sont alors bien réelles, puisqu'on est prêt à jouer pour elles. Le coût S_n est ici la "surprise" et la quantité d'information la "surprise moyenne".

Par contre, dans le cas où le message peut continuer à arriver, l'estimation de H peut être comparée à celle que donne la statistique et se révéler inexacte ou exacte.

Dans le premier cas, l'estimation des p_n par les q_n peut donner l'illusion d'une perte d'information ou d'un excès d'information tout-à-fait fictif. Mais la probabilité p que le message p_n soit une fluctuation de q_n diminue progressivement ; ce p (ou d'autres) peut servir de "poids" pour définir l'"information équivalente" = Kp + H (1 - p).

Pour un très long message, p = 0 et l'information apportée par le message est bien H, qui peut se révéler très différent de K.

5.4.4 - INFORMATION ET CAPACITE

Bien que p_n puisse ainsi servir d'estimation de p_n, il faut distinguer absolument les deux notions du point de vue conceptuel. De même, la notion d'information $H - \sum p_n \log p_n$, associée au message, c'est-à-dire à un processus stochastique, diffère de la notion de capacité $K = - \sum q_n \log q_n$ associée à une méthode de codage.

Cependant, on les confond souvent dans des raisonnements du type suivant : On a un appareil réel fixe, on calcule le nombre d'alternatives que cet appareil transmet pour définir une image, et on parle de l'"information de l'appareil", c'est-à-dire de la méthode de codage. Ceci n'a aucun sens : on peut tout au plus définir la redondance d'un appareil pour un message donné, ou la redondance moyenne, ou maximum. Pour définir une information à partir d'un appareil, il faut considérer non pas un seul, mais tous les appareils possibles pour un message donné, et prendre le minimum de $\sum p_n \log q_n$ sur cet ensemble d'appareils.

$\sum q_n$ loc q_n caractérise excellement la complication fonctionnelle de l'instrument, mais pas en tant qu'entropie : on ne doit pas la poser comme terme d'un bilan entropique.

La considération de capacité subsiste dant tout mécanisme pouvant être interprété comme (D).

5.4.5 - PARADOXE DE L. BRILLOUIN

La confusion entre H et K peut mener à des paradoxes tout-à fait artificiels, comme celui de L. Brillouin (1949, 1950) qui va être discuté ici.

Soit un texte écrit ou un film :

(D) consiste simplement à envoyer un faisceau de lumière : en principe, le film joue le rôle de filtre inerte décomposant la lumière en partie réfléchie (ou transmise) et partie absorbée et diffusée.

Il y a entre (D) et M une "adaptation mécanique" poussée, mais très mauvaise adaptation thermodynamique, se traduisant par l'absence de relation entre les propriétés physiques du véhicule et le message transmis.

Or Brillouin appelle hâtivement "négentropie du message" ce qui n'est que la partie utile de l'augmentation d'entropie du faisceau incident : dans ces conditions, si 1000 personnes lisent un livre, son information ("négentropie") est multipliée par 1000 ! Mais où est alors le 2ème principe de la thermodynamique ? Brillouin le retrouve en distinguant arbitrairement entre deux espèces d'entropie, dont une seule suivrait le principe, sans indiquer d'ailleurs comment identifier en général ces deux parties.

Il est clair que l'erreur a consisté à poser des quantités d'information et des capacités dans un bilan entropique, en omettant précisément la vraie entropie physique du faisceau. Dans chacune des opérations de lecture, celle-ci augmente de quantités très supérieures à l'information transmise, tandis que le livre-message reste intact (du moins en première approximation, car l'absorption ou diffusion a pour effet de détruire le texte, qu'on le lise ou non : en somme un peu d'entropie provient du message). En d'autres termes, l'erreur tient à ce que l'on ne s'est pas strictement tenu à un seul système, de préférence clos, dans toute la suite des raisonnements mais que l'on a confondu objet, instrument et résultat de l'expérience. On ajoute et retranche des variations d'entropie appartenant à des systèmes différents. Dans ces conditions, il est inutile d'espérer que le deuxième principe s'applique sans tiraillements.

Pour conclure, on pourrait dire que le paradoxe de Brillouin tient à une extension à la théorie de l'information de la théorie des sens d'Empédocle. Celui-ci affirme que, pour percevoir de la lumière ou du son, l'on doit avoir dans l'organe de sens de la lumière ou du son de même nature physique. Brillouin identifie à priori quantité d'information et capacité en les mettant des deux côtés d'un bilan. Dans les deux cas, cette identification n'est pas physique, mais métaphysique, et dans les deux cas on ne peut introduire de sens physique qu'en remplaçant l'hypothèse d'identité par celle de l'adaptation du message et du récepteur. Le schéma fonctionnel correspondant suit pas à pas les lois de la physique et ne comporte que des opérations qu'elle autorise.

CHAPITRE 6

CODAGE SÉQUENTIEL

6.1 - POSITIONS DES PROBLÈMES

6.1.1 - PROBLEMES MICROSCOPIQUES

Nous avons indiqué que l'information sélective est la seule à considérer dans le cas du message discret, tandis que le message physique est toujours continu. Il a fallu les Chapitres 4 et 5 pour nous dégager des difficultés que cela entraîne, mais nous sommes maintenant ramenés à un problème purement discret, que nous allons d'abord considérer du point de vue microscopique.

Ceci signifie encore que le message est considéré dans ses éléments : c'est une suite de tirages au sort entre R mots M_n ($1 \leqslant n \leqslant R$) (Cf. § 5.2 au sujet du terme "mot") dont les probabilités p caractérisent le message. Ce point de vue statistique dans le problème de transmission de l'information est surtout dû à Kolmogoroff et Wiener.

Le point de vue macroscopique sera abordé au Chapitre 7. Le message sera alors considéré comme un tout et caractérisé par un certain nombre de variables d'état.

Quant au codage, c'est une règle de correspondance entre mots et suites de symboles que l'on peut envoyer sur la ligne.

6.1.2 - LES QUATRE TYPES DE PROBLEMES DE CODAGE

Selon les caractères de la correspondance entre message et codage, et en tenant compte des erreurs possibles dans la transmission et de celles admissibles à la réception, on obtient respectivement quatre types de problèmes :

- A/ Un codage par message et réciproquement. Ex : problème de Shannon, Fano, du meilleur de tous les codages.
- B/ Plusieurs codages par message : codes surabondants déceleurs ou correcteurs d'erreur. Il faut adapter les surabondances initiales aux ambigüités futures.
- C/ Plusieurs messages par codage : codes ambigüs. Ce sont des transducteurs irréversibles.
- D/ Codes à la fois surabondants et ambigüs.

A chaque fois, on a un couple de problèmes direct-inverse, selon que c'est le message ou le codage qui est donné.

Nous nous poserons le problème de codage optimum mot-par-mot, avec top de synchronisation après chaque mot.

D'après les résultats de Shannon, un tel codage est nécessairement surabondant. Il a pour effet que les erreurs dues à une mauvaise transmission d'un symbole élémentaire se bornent à rendre inintelligible un seul mot, et non tous le message (mais il peut aussi avoir un effet correcteur plus prononcé : § 6.3.5).

Le premier problème, d'adaptation directe, se pose lorsque le message est donné, ainsi que les symboles élémentaires du codage, et qu'il s'agit de trouver la meilleure loi de correspondance entre mots et groupes de symboles.

Le deuxième problème, qui est inverse du précédent, se pose lorsqu'il s'agit de trouver le message qui utilise au mieux une loi de correspondance connue, déterminée par un problème direct relatif à une classe de messages, dans un cas où la solution ne dépend pas du message particulier mais de la classe. Le problème inverse détermine alors le meilleur message dans cette classe. Il ne semble avoir été abordé nulle part jusqu'ici. Il fera l'objet du § 6.3.

Pour le poser, il faudra posséder la notion d'information sélective, qui s'introduit de la façon la plus élégante par le "problème de Shannon". Celui-ci sera donc esquissé au § 6.2 après le problème direct limité nécessaire au problème inverse. Le problème de Shannon est de ceux pour lesquels on peut procéder effectivement à l'optimisation à l'intérieur d'une "classe" qu'appelle la définition même des problèmes inverses. Ceci veut dire que l'expression $\Sigma p_n \log p_n$ de l'information sélective peut s'obtenir non seulement comme la solution d'un groupe d'axiomes appropriée au cas considéré, mais aussi par le passage à la limite à partir de codages effectifs. (Notons que c'est précisément cet exemple qui a été à l'origine de toute la théorie générale de la première partie).

6.1.3 - REMARQUES MARGINALES SUR LES PROBLEMES DES CODES AMBIGUS

Comme ils consistent à introduire du bruit actuel à bon escient, ces problèmes sont exactement inverses du problème de l'élimination du bruit potentiel dont Shannon a montré qu'il est théoriquement possible, mais non réalisable. L'intérêt du codage ambigü vient de la rapidité de la variation de l'équivocation et de la complication de l'appareillage en fonction de l'erreur effective ou de l'erreur admissible. Par suite, dans tous les cas où une erreur, même faible, est admissible, on peut en profiter pour : soit diminuer le coût de la transmission par unité d'information effectivement utilisable, c'est-à-dire diminuer la capacité de la ligne; soit simplifier les appareils; soit faire les deux à la fois.

La deuxième alternative est très intéressante, car elle revient à court-circuiter le code "parfait" exact optimum, en sautant directement du code non ambigü mais de mauvais rendement, à un code ambigü de bon rendement. Les deux codes qu'on relie ainsi sont tous deux simples, tandis que le code parfait aurait été en général très compliqué et peut-être non réalisable.

L'équivocation étant une fonction du bruit qui laisse sa forme exacte assez indéterminée, cette forme peut être construite dans chaque cas de la façon la plus appropriée au but cherché. Pour cela, on peut, soit chercher directement le bruit le plus approprié, soit chercher inversement à identifier le bruit parmi une famille aussi riche que possible de bruits construits à priori. (On voit la liaison étroite entre le problème de la construction du bruit et celui de l'adaptation à une stratégie, lorsque celle-ci est définie par une mécanique dont on ne connaît pas le fonctionnement).

Exemple des cartes perforées; de F lignes de G points chacune. La capacité de la carte est FG et elle représente à la fois le maximum de l'information qu'elle peut porter, et le coût du codage sur cette carte de n'importe quelle information, même inférieure.

L'existence des lignes rend la carte appropriée au codage ligne par ligne. Or celui-ci n'est économique que si chacune des dichotomies naturelles qui définissent le message exige G unités d'information binaire.

En fait les dichotomies naturelles sont simples : chacune traduit la présence ou l'absence d'un parmi un grand nombre de caractères indépendants. Mais ceci exige G = 1 et F grand, et de plus ces dichotomies simples vont entre éléments de probabilités très différentes, donc le codage exact, simple et décomposable en éléments d'un seul symbole binaire, est très redondant. Pour éviter cette redondance, l'on devrait construire des dichotomies artificielles multiples, avec alternatives équiprobables; ce qui rendrait l'encodage et le décodage très compliqués.

Par suite, l'acceptation d'une ambigüité suffisante permet non seulement d'augmenter l'information maximum transmissible, mais aussi d'utiliser beaucoup mieux ce maximum avec des moyens très simples.

Système Zator. Cf. Mooers (1950). C'est un code ambigü non séquentiel, qui a été imaginé en vue de recherches bibliographiques, c'est-à-dire de l'isolement rapide de tous ceux des objets d'un ensemble donné, qui possèdent une certaine combinaison de caractères. La présence erronée d'objets additionnels ne présente pas d'inconvénient.

A chacun des caractères indépendants, on associe un code de N symboles choisis au hasard parmi les C_F^N codes possibles. Pour encoder, on superpose sur une même carte les codes de tous les caractères que possède l'objet représenté par le message. Pour décoder, on vérifie successivement si l'objet possède les divers caractères possibles.

Comme les codes ont en général des points communs, la réunion des points appartenant aux caractères réels contient toujours des codes appartenant à des caractères que l'objet ne possède pas. Tout se passe comme si l'on avait un top "flottant", et que l'erreur sur chaque message consistait à le transmettre avec le top repoussé plus loin et des symboles additionnels au bout, le rendant beaucoup plus explicite qu'il n'est en réalité (si l'on tient compte d'éléments "sémantiques" les caractères ne

sont pas absolument indépendants, et le message reçu peut être absurde.) Par suite, le message reçu est toujours plus déterminé que le message envoyé; en d'autres termes, le nombre de messages reçu possibles est plus petit que le nombre de messages envoyés.

On a en même temps diminution de l'information apparente et équivocation. Le bruit est toujours dans un même sens : le nombre de caractères attribués est supérieur à la réalité.

Codes ambigüs considérés comme étapes.

Le problème bibliographique auquel le code Zator est approprié tire le plus clair de son intérêt théorique de son application possible au problème de la mémoire, considérée comme réservoir d'où il faut tirer le plus rapidement et économiquement possible le code précis qui correspond à la description, souvent imprécise, que l'on possède. Ce problème du présélecteur de code n'a pas été beaucoup étudié pour lui-même, et il est évident que la théorie des communications habituelle ne peut nous permettre de réduire les dimensions de la mémoire, puisque cette réduction exigerait un code plus ou moins optimum, et qui reposerait le même problème.

La réduction de dimension permise par Zator provient d'une division des opérations en plusieurs étapes :

La 1ère exige un code à faible mémoire pour les caractères standard.

La 2ème permet de sortir très rapidement un nombre réduit de fiches parmi lesquelles se trouvent des fiches inutiles mais aussi toutes les fiches cherchées. Au codage ambigü se juxtapose un codage précis constitué par le titre complet et de plus, dans le cas de la mémoire, le code optimum cherché. En somme, après réception du message ambigü, on sollicite l'envoi d'informations complémentaires qui lèvent l'ambigüité et en même temps augmentent l'information apparente. On retombe donc en fait sur un transducteur réversible, dont certaines opérations auront été court-circuitées sans danger.

6.2 - PROBLÈME DIRECT
ADAPTATION D'UN CODAGE ARITHMÉTIQUE AU MESSAGE DISCRET

6.2.1 - MOTS INANALYSABLES

Soit R le nombre de mots différents. On ne peut les transmettre immédiatement sans les analyser que si l'on dispose d'une ligne ayant R états différents E_g. Le codage consiste alors uniquement à donner une correspondance un à un entre les mots classés par fréquences décroissantes, et les états classés par coûts croissants. L'étude s'arrête là.

6.2.2 - LE CODAGE CANONIQUE MOT-PAR-MOT

En général, le nombre d'états de la ligne est très inférieur au nombre de mots. On doit alors représenter chaque mot M_n par une suite de "lettres" et faire se correspondre "lettres" et états de la ligne de transmission.

Voyons d'abord comment se présente le problème si l'on veut transmettre les mots M_n mot par mot, sans les regrouper avant la

transmission. Alors, le message se présentera comme une suite de notes écrites sur une portée à q + 1 lignes. La ligne E_0 sera réservée aux tops de fin de mot. Les lignes E_g $(1 \leqslant g \leqslant q)$ seront utilisées pour la représentation des mots.

L'envoi des signaux E_g $(0 \leqslant g \leqslant q)$ ne peut pas se faire sans coût C_g. Ceci résulte du § 5.2.3. Ici nous nous désintéressont des raisons de ce coût, mais le supposerons mesurable avec une unité arbitraire. Au § 6.2.5 s'introduira une unité de coût naturelle, fonction d'état du codage.

Il y aura adaptation physique de la stratégie de codage aux coûts des symboles et aux probabilités des mots, lorsque le coût moyen sera minimum pour les p_n donnés. Le codage correspondant sera dit canonique.

Pour cela, il faut encore que le classement des mots par probabilités décroissantes corresponde à celui par coûts croissants des groupes de symboles E_g qui représenteront ces mots. Mais ce dernier classement n'est plus donné ici, comme il n'était au § 6.2.1, et il pose un problème d'algèbre dont la solution est donnée dans l'Appendice § 6.2.7.

On trouve que le coût du $n^{ème}$ groupe de signes par ordre de coûts croissants peut en général être mis sous la forme :

$$j_n = \left[j_0 + \log_M (n + m) \right]$$

où $[x]$ est le plus petit entier supérieur à x.

Les constantes M, j_0, m dépendent de l'ensemble des nombres C_g. Nous les appellerons les "variables d'état" du codage car les C_g eux-mêmes n'auront plus jamais à intervenir directement dans la théorie (si toutefois le nombre des messages est suffisant : dans le cas de la langue, ceci exclura les phonèmes, trop peu nombreux pour que l'influence des détails du codage cesse d'être sensible). Les variables d'état jouent pour les lignes discrètes, le même rôle que le rapport signal/bruit et la largeur de bande jouent pour les lignes continues.

Mentionnons simplement que M est la plus grande racine réelle de $\sum_1^q M^{-C_g} = 1$. Exemples :

1°: cas simple : $C_g = 1$ $M = q$, $m = q/(q-1)^{-1} > 1$

2°: cas quasi-simple : $C_g = g$, $q = \infty$ $M = 2$, $m = 1$

3°: cas semi-simples : tous les autres cas. Le codage est beaucoup plus compliqué, mais présente en revanche l'avantage que M peut devenir très proche de 1 . Alors la courbe quantifiée $j_n = f(n)$ s'écarte beaucoup moins d'une courbe continue que dans le cas du codage simple ou quasi-simple.

Un autre avantage du codage quasi-simple, et de tous les autres codages où les coûts ont des valeurs très différentes, est que les signaux élémentaires "chers" ne sont nécessaires que pour des signaux rares, c'est-à-dire que les moyens de codage augmentent en même temps que le vocabulaire qu'il doit exprimer.

On peut rapprocher ce dernier fait de l'exemple des "routines" dans la computation mécanique, c'est-à-dire des ensembles d'opérations réglées à l'avance, que l'on "compose" pour obtenir

l'opération résultante. Le nombre des "routines" augmente également avec la variété des opérations à effectuer.

Délai dans la reconnaissance et coût en codage quasi-simple.

R.W. Hick (1951) a mesuré le temps moyen que met un sujet à faire un mouvement simple signifiant qu'il a identifié un stimulus simple. Il a fait varier le nombre de stimulus possibles équiprobables. Le temps de réaction a été trouvé être :

$$t_r = 0{,}27 \log_e \ (n + 1)$$

L'interprétation de Hick est la suivante : " t_r est proportionnel au minimum de quantité d'information qui doit être extraite du signal pour l'identifier, en supposant que le cerveau fonctionne comme si l'absence de signal était un signal de probabilité égale à celle de chacun des signaux réels. Le processus d'analyse est interprété comme succession d'opérations d'identification". Il est clair que la seule justification de l'équiprobabilité est que le résultat est correct.

Il nous semble plus raisonnable d'interpréter ce coefficient comme le "m" de notre théorie. Si sa valeur est exactement égale à 1, on a un codage quasi-simple. Une valeur seulement voisine de 1 pourrait signifier codage simple avec q élevé ou bien d'autres combinaisons. Le fait que $t_r = 0$ pour $n = 0$ n'a d'ailleurs bien évidemment rien d'étonnant, puisque m a dès l'abord été défini de façon à rendre nul le nombre de signaux de coût nul.

Il se peut que les états E_g dont on dispose, ne puissent être obtenus que par synthèse à partir d'un certain nombre de fibres plus élémentaires utilisées simultanément, dont chacune a moins de q états. On doit alors faire successivement l'analyse dans le temps et l'espace, ce qui introduit deux ordres successifs de messages d'analyse avant d'arriver au codage. On pourrait même avoir à introduire des ordres plus poussés d'analyse, jusqu'à ce qu'un nombre, petit ou grand, d'éléments différents aura été ramené à un grand nombre d'éléments identiques; en général, on ne pourra, ni ne voudra, aller au-delà de messages binaires.

(L'ordre des analyses : temps puis espace est préférable à l'ordre inverse, car les mécanismes étant donnés, il permet de les utiliser au maximum, en économisant le temps. Cet ordre n'est cependant pas bien défini dans le cas où les symboles des lignes en question sont en réalité des suites dans le temps des symboles d'une seule ligne. Mais les deux analyses, étant alors indépendantes, peuvent être considérées comme faites dans des "directions" orthogonales dont la définition est de pure commodité).

En considérant le procédé de transmission ci-dessus, on englobe tous les codages arithmétiques jusqu'ici envisagés.

6.2.3 - LE CODAGE PAR LARGES BLOCS DE SHANNON

Nous avons trouvé ainsi que le codage canonique ne dépend pas des probabilités p_n elles-mêmes, mais seulement des propriétés des symboles transmissibles.

Le coût moyen par mot est $\sum p_n j_n$. On n'est pas maître de p_n pour diminuer ce coût moyen, sauf si l'on accepte de changer la définition des mots, en regroupant ceux-ci.

C'est là une idée fondamentale, due à Shannon (1948), et qui conduit à la définition de l'information sélective, dont nous allons donner ici une justification heuristique. Changeant de point de vue, appelons "nouvel élément" l'ensemble de deux "anciens éléments" $M_{mn} = M_n + M_m$. D'après le théorème de probabilités composées, sa probabilité sera $p_{mn} = p_m\, p_n$.

Augmentons ainsi progressivement le nombre N d'éléments anciens dans un élément nouveau. D'après la loi des grands nombres, l'élément M_n s'y retrouvera avec une fréquence égale à Np_n, sauf éléments exceptionnels extrêmement peu probables, que nous négligerons.

Par suite, la probabilité de sortie d'un élément nouveau quelconque sera indépendante de cet élément et égale à :

$$P = \text{produit de tous les } (p_n)^{Np_n}$$
$$= (\text{produit de tous les } p_n^{p_n})^N = M^{NH_M}$$

en posant :

$$H_M = -\sum p_n \log_M p_n$$

Inversement, M^{NH_M} sera le nombre des éléments équiprobables de fréquence appréciable.

Pour signaler le résultat qu'il a obtenu, l'observateur du tirage peut se contenter d'envoyer le numéro de l'élément nouveau dans un classement arbitraire. Ce numéro spécifiera entièrement ce qui lui a appris ma suite de tirage au sort, puisqu'il permettra de la reconstituer. Etant donné l'arbitraire du classement, on doit calculer la longueur moyenne du numéro qu'il faut envoyer. Ce sera, en négligeant m et j_0 :

$$\frac{1}{M^{HN}} \sum \log_M n = NH_M$$

soit H_M symboles par élément de la définition d'abord adoptée.

Le raisonnement qui précède sera d'autant plus rigoureux, que sera plus grand le nombre d'éléments primitifs réunis dans un seul élément nouveau. Ce codage qui fait intervenir le nombre strictement minimum d'éléments ne peut donc s'obtenir que si l'on regroupe les mots en blocs infiniment longs. Shannon a par ailleurs montré que ce nombre minimum de symboles H_M est effectivement la seule expression qui satisfasse à tous les axiomes désirés d'une quantité d'information, dans le contexte présent : celui d'une transmission par symboles "séquentiels" (nous avons vu en détail aux Chap. 2 et 4 qu'à une "stratégie" de transmission autre que séquentielle, correspondent d'autres solutions du système d'axiomes de l'information). La définition de la quantité d'information est fondamentale. Cependant, les théorèmes de Shannon sur le problème direct ne peuvent jouer qu'un rôle comparable aux théorèmes de Wiener sur la réalisabilité des filtres ayant certaines caractéristiques, comme limites de combinaisons d'éléments réalisables. C'est dire qu'ils ne nécessitent aucune utilisation des détails de la structure de la loi p_n, mais en revanche ne permettent pas d'utiliser ces détails dans l'étude

de la réalisabilité exacte du codage optimum, ni dans l'étude physique précise des cas individuels (cf. § 6.4).

6.2.5 - COUT D'ENVOI DE L'UNITE D'INFORMATION NEPERIENNE

Lorsque ce coût est minimum, les probabilités r_g des signaux E_g de coût C_g sont données par :

$$\frac{C}{H} = \frac{\Sigma r_g C_g}{-\Sigma r_g \log_e r_g} \quad \text{minimum, soit}$$

$$H_e \Sigma \Delta r_g C_g - C \Sigma \Delta r_g \log_e r_g - H^2 \Sigma \Delta r_g = 0$$

ce qui donne :

$$r_g = Pe^{-BC} g$$

$$C = \Sigma r_g C_g$$

$$H_e = B \Sigma r_g C_g - \log_e P$$

Le coefficient B est déterminé par :

$$B = \frac{H}{C} = B - \frac{\log P}{C} \qquad \text{soit } P = \Sigma_1^q e^{-BC_g} = 1$$

Par suite, le critère suffit, même pour déterminer le coût moyen et la quantité d'information. Posons :

$$e^{-B} = M \qquad \Sigma_1^q M^{-C_g} = 1 \qquad \boxed{r_g = M^{-C_g}}$$

$$(1 \leqslant g \leqslant q) \qquad C = \Sigma r_g C_g \qquad H_e = (\Sigma r_g C_g) \log_e M$$

$$\text{donc} \quad \boxed{C/H_e = (\log_e M)^{-1}}$$

Le nombre M est le même que celui du § 6.2.2.

Revenons maintenant au résultat du § 6.2.4. Nous avons vu que le coût moyen est H_M dans le codage à la Shannon. En d'autres termes $C/H_e = (\log_e M)^{-1}$. Donc le codage à la Shannon réalise la meilleure répartition de l'information entre symboles.

Par ailleurs, le résultat sur C/H_e permet de lever l'indétermination sur la valeur absolue du coût. On peut par exemple, convenir de prendre une unité népérienne telle que $\log_e M = 1$, ou $M = e$. Alors une unité de coût népérien transporte une unité d'information népérienne.

6.2.6 - THERMODYNAMIQUE DISCRETE

La Thermodynamique du Chap. 4 était l'étude des relations entre les concepts d'information sélectif et fishérien, et les propriétés physiques du signal. Au chapitre 5, nous avons montré que les propriétés du signal pouvaient être entièrement représentées par des coûts. Par ailleurs, l'information fishérienne est ici sans objet. Cependant, il reste intéressant d'appeler Thermodynamique discrète l'étude des relations entre coût et information sélective.

Cette étude est beaucoup plus simple que celle du Chapitre 4. Tout d'abord le signal discret est d'entropie nulle, c'est-à-dire qu'aucune de ses propriétés n'empêche d'atteindre le maxi-

mum d'information concevable. Quant à la température, c'était un coefficient d'équivalence de maximum d'information et d'énergie, qui, étant de dimensions différentes, ne pouvaient être fournis d'utilisés correspondantes. Ici, au contraire, si on fait se correspondre des unités népériennes, il y a égalité entre coût et information potentielle. Si les unités ne se correspondent pas, le coefficient d'équivalence est $(\log_e M)^{-1}$, mais comme il n'est pas intrinsèque, nous n'appellerons pas ce nombre température. D'ailleurs le rôle central de la température thermodynamique est plutôt pris par un autre paramètre, qui s'introduira plus loin (§ 6.3). Pour tout codage autre que le codage concret de Shannon, on peut aussi définir un rendement de la stratégie utilisée, ainsi que la redondance ou moins le rendement.

6.2.7 - SOLUTION DU PROBLEME D'ALGEBRE DU § 6.2.2.

Appelons σ le plus grand des C_g, lorsque tous ces nombres sont des entiers.

Etude de M(j).

Soit M(j) le nombre de messages de coût total égal à j. Il est racine de l'équation homogène aux différences finies

$$- M(j) + \Sigma M\ (j-C_g) = 0$$

Cette racine est de la forme

$$M(j) = \Sigma_1^{\sigma} A_s M_s^j$$

Ces σ nombres M_s^{-1} sont les racines de l'équation caractéristique $\Sigma z^{C_g} - 1 = 0$, à coefficients entiers non négatifs sauf le terme constant -1. Les σ coefficients A_s sont déterminés par les σ équations linéaires résultant des conditions aux limites suivantes :

$M(j) = 1$ pour $j = 0$ (c'est le "message" "rien du tout" qui n'a pas de réalité physique) et $M(j) = 0$ pour $-\sigma + 1 \leqslant j < 0$.

Le M(j) résultant de ces conditions aux limites ne sera plus nul pour $j \leqslant -\sigma$; mais cette partie de M(j) sera sans intérêt physique, ainsi d'ailleurs que celle de $-\sigma + 1$ à 0.

1er exemple de M(j) : Code simple q - adique

$C_g = 1$; $qM^{-1} = 1$; $M = q$; $M(j) = q^j$ $(j > 0)$.

C'est le cas où tous les symboles sont de même coût.

Variante : Si, dans l'exemple précédent, on change d'unité et C_g devient h, alors $M_s = qW\binom{h}{s}$ où $W\binom{h}{s}$ est la s^e racine h^e de l'unité. Les A_s doivent être tels que $M(j) = 0$ sauf pour les multiples de h où $M(j) \leqslant (q)^{j/h} = (q^{1/h})^j$. Lorsque les C_g sont différents mais tous très voisins les uns des autres et de h, les plages entre multiples de h sont à peu près aussi peuplées que les multiples eux-mêmes, tout au moins à partir d'un n assez grand, on a approximativement :

$$M\ (j) = \frac{1}{h}\ (q^{1/h})^j$$

Un système dyadique ne peut être compatible avec la transmission de l'information par mots, car il ne resterait aucun signe pour les tops. Il est cependant possible de construire des

signaux complexes à partir de 0 et 1. L'exemple suivant semble physiologiquement le plus réaliste :

2ème exemple : Code quasi-simple

$$C_g = g; \quad M^{-g} = M^{-1} \frac{1 - M^{-q}}{1 - M^{-1}} = 1$$

Si $M \neq 1$, ceci donne pour $M^{-1} = Z$ l'équation

$$Z - Z^{q+1} = 1 - Z, \quad Z^{q+1} - 2Z + 1 = 0$$

(dont la racine $Z = 1$ est illusoire, donc qui est bien de degré q).

Pour $q = \infty$, une seule racine $M = 2$, $M(j) = 2^{j-1}$. Il serait bien entendu anormal d'utiliser un nombre infini de lettres pour coder un nombre de mots fini. Mais l'absence de lettres de coût q ne se fait sentir que pour n supérieur au n_o tel que $j_{n_o} = q$. En deçà de ce n_o les autres racines de l'équation caractéristique n'ont aucune influence.

3ème exemple. $C_g = c + gd$ $(0 \leqslant g \leqslant q)$, c, d premiers entre eux. Alors :

$$M^{-c} \left(1 - M^{-d(q+1)}\right) = 1 - M^{-q}).$$

Pour $q = \infty$, $M^{-c} = 1 - M^{-d}$.

On a toujours pour M(j) une somme des M . Le nombre de ses termes étant le plus grand des C_g devrait être ici ∞, mais dans ce cas particulier la somme ne s'étend que jusqu'à sup (c,d) termes.

L'alphabet Morse appartient à cet exemple avec $q = 2$. En effet le coût de l'intervalle entre lettres = coût du signal 0; le coût du point = coût du signal 1 + signal 0; le coût du trait = coût de 2 signaux 1 + signal 0.

4ème exemple : $C_g = C + g^2 d$, c, d premiers entre eux. $(0 \leqslant q \leqslant q)$. Il faut :

$$M^{-c-g^2-d} = 1; \text{ Si } q = \infty, \ M > 1$$

$$1 = M^{-c} \sum e^{-g^2 d (\log M)} \sim M^{-c} \sqrt{\frac{\pi}{4d \log M}}$$

$$\log M = \frac{\pi}{4d} M^{-2c}; \text{ d'où } x \log x = \frac{\pi C}{2d} \text{ avec } x = M^{2c}.$$

Pour que M puisse être $> e$, $\log_e M > 1$, il faudrait que $\frac{\pi}{4d} M^{-2c}$, qui serait $< \frac{\pi}{4de^{2c}}$, soit aussi > 1, ce qui est impossible même avec $d = c = 1$; donc $M < e$. On obtient une limite inférieure de M en remplaçant les deux courbes log M et M^{-2c} par leurs tangentes en $M = 1$. Alors

$$M-1 = \frac{\pi}{4d} (1 - 2cM) = \frac{\pi}{4d} - \frac{c}{2d} M$$

$$M = \left(1 + \frac{\pi}{4d}\right) \left(1 + \frac{c}{2d}\right)^{-1}$$

La valeur exacte de M diminue quand, c/d variant peu, d et c augmentent.

Etude de N(j).-

Soit ensuite N(j) le nombre de messages de coût total plus petit que j ("rien du tout" étant exclu).

N(j) est, bien entendu, solution de la même équation que M(j), mais pourvue d'un deuxième membre inconnu. Calculons donc N(j) en sommant les M(j)

$$N(j) = \Sigma_1^j M(i) = \Sigma_1^j A_s \frac{M_s}{M_s - 1} (M_s^j - M_s) + \Sigma A_s M_s$$

$$= \Sigma B_s M_s^j - m$$

où $$m = \Sigma A_s \frac{M_s^2}{M_s - 1} - \Sigma A_s M_s = \Sigma \frac{A_s M_s}{M_s - 1}$$

C'est une fonction symétrique, donc réelle, de l'équation caractéristique; elle est par suite exprimable en fonction algébrique des coefficients de cette équation.

1er exemple :

$$N(j) = \frac{q^{j+1}}{q-1} - \frac{q}{q-1} \; ; \; m = \frac{q}{q-1} > 1$$

2ème exemple :

$$q = \infty : N(j) = 2^j - 1 \; ; \quad m = 1.$$

m est le même que si l'on avait une infinité de symboles de coûts égaux. Pour M : même résultat que si l'on avait deux symboles de coûts égaux.

Solution approchée dans le cas général (semi-simple).

L'équation aux différences finies n'est pas exactement soluble.

Soit M la racine réelle de l'équation caractéristique. Par application de la règle de Descartes, $M > 1$. En première approximation, le coût du n^e message par coûts croissants sera $j_n = \log_M n$.).

En deuxième approximation :

$$n = \frac{M}{M - 1} M^j + \Sigma' \frac{A_s M_s}{M_s - 1} M^{\log_M M_s} - m$$

où Σ' est la somme qui omet la racine M.

Nous dirons que le système de coût est stable si

$$\mathrm{Re}\,(\log_n M_s) < 0$$

Alors Σ' 0 avec 1/n et $j_n = [j_0 + \log_M (n+m)]$. Les cas simple et quasi-simple donnent des exemples de stabilité parfaite.

Dans les cas très stables, les A_s sont petits pour $s > 1$ et les M_s sont grands. Par suite, $n_0 \sim \frac{M}{M - 1}$. Dans les cas où il y

a des racines réelles négatives < - 1, il n'y a pas stabilité, mais ceci influe beaucoup moins sur N(j) qui reste assez stable, que sur M(j) qui subit des oscillations importantes.

6.3 - PROBLÈME INVERSE ADAPTATION DU MESSAGE AU CODAGE CANONIQUE

6.3.1 - LA CORRELATION

Un des points les plus intéressants de la théorie de Shannon est l'introduction de la corrélation. Mais en fait, c'est sans utilité pratique dans la majorité des cas, car peu de mécanismes peuvent en tenir compte. Le message doit donc être découpé en messages individuels. Plus courts sont ces messages, plus souple est l'appareil de transmission (et plus court le délai) mais plus médiocre en général le rendement. Le découpage comporte donc un problème de compromis entre ces deux exigences opposées. Nous chercherons à répartir au mieux l'information entre mots en cherchant quelle doit être la statistique de ceux-ci pour que le message soit le mieux adapté à la transmission sur une ligne arithmétique, représentée par ses variables d'état. C'est là un problème inverse de codage.

Les critères de l'adaptation seront aussi généraux que possible : nous les avons choisis indépendamment les uns des autres, mais les lois auxqlelles ils mènent se révèleront former des sous-familles se chevauchant mutuellement d'une même famille canonique. Ce résultat établira des liaisons à posteriori entre les critères utilisés.

6.3.2 - HYPOTHESES FONDAMENTALES

Les deux hypothèses du problème inverse sont donc les suivantes :

1°) Les mots sont définis à priori, leur corrélation négligée et le codage fait mot par mot. (Si leur définition est arbitraire et se révèle incorrecte, le principe d'adaptation pourra la rectifier à posteriori. § 6.3.4).

2°) Des tops séparent les groupes de lettres représentant deux mots différents. Ces tops permettent d'utiliser pour deux mots différents des codages dont l'un est identique à l'autre suivi de quelques nouveaux signes. Ceci était impossible dans les codages à la Shannon, car là, aucun fragment du codage d'un mot ne devait pouvoir représenter d'autre mot. Le résultat était que Shannon pouvait se passer de tops. Mais comme, en pratique, en leur absence, une seule erreur détruirait tout le message, on aurait, quand même, été amené à les envoyer (§ 6.3.5).

Leur considération dès le début n'est donc, comme la première hypothèse ci-dessus, que tout-à-fait réaliste, en opposition avec les hypothèses opposées de Shannon, qui, elles, sont peu réalistes. Les tops seront de plus indispensables pour concevoir l'adaptation probabiliste (§ 6.3.3).

Malgré leur apparence inoffensive, ces deux hypothèses mo-

difient beaucoup la position du problème, et permettent d'aboutir à une loi canonique (§ 6.3.3.). De la théorie de Shannon, il ne restera guère que la définition précise de certaines des qualités "globales" rattachées au signal, dont la plus importante est la "quantité d'information". Il ne semble pas que l'on puisse, dans l'étude de la synthèse, progresser beaucoup au-delà des résultats de Shannon, avant d'avoir effectué une analyse concrète des divers codages, ce qui équivaut à la synthèse artificielle des messages.

TYPES DE CRITERES D'ADAPTATION

Le codage adopté sera toujours le codage canonique du § 6.2.2 qui est le plus économique une fois les p_n donnés (si l'on transmet mot par mot avec tops). On pourra donc changer les règles de la répartition de l'information entre mots sans que ce codage soit à changer.

Les critères d'adaptation sont de deux types possibles : dans le critère du premier type, le coût du top n'intervient pas à priori; dans les trois critères du deuxième type, le coût du top intervient explicitement.

6.3.3 - ADAPTATION DE PREMIER TYPE (ECONOMIQUE)

Dans le premier critère, nous supposeront que l'information à transmettre par mot est fixée à l'avance, mais pas les p_n individuels. La prévoyance de l'émetteur est alors réduite à un seul mot (émetteur imprévoyant). Il est cependant libre de la façon de répartir l'information parmi les mots qu'il possède, dont le nombre est R (Rang maximum). De ce point de vue, la meilleure distribution des fréquences des mots sera celle pour laquelle le coût moyen du mot $\Sigma\, p_n\, j_n$ sera minimum, étant donnés la quantité d'information par mot $-\Sigma p_n \log_M p_n$ et le "nombre potentiel" de mots R.

La méthode des multiplicateurs de Lagrange permet de déduire de ce principe variationnel global, une loi canonique locale

$$p_n = Pe^{-B' j_n}$$

Pour pouvoir tenir compte du fait que j_n est essentiellement un logarithme de base M, on écrira plutôt :

$$p_n = P\, M^{-B j_n} = P \left[n + m\right]_M^{-B}$$

où $[x]_M$ est le nombre dont le logarithme est égal au plus petit entier supérieur à $\log_M x + j_0$.

Nous venons ainsi d'introduire un nouveau paramètre d'état; B, qui peut prendre n'importe quelle valeur positive et se trouve dans chaque cas déterminé par la condition de donner la valeur correcte de H (cf. § 6.4.2).

L'inverse de B, que nous écrirons θ et appellerons température informationnelle, jouera le rôle de variable d'état fondamentale. Ce n'est pas un rendement, une telle notion ne pouvant même pas être définie en l'absence de tout renseignement sur le coût du top, comme c'est le cas ici.

RAISON POUR LAQUELLE ON ABOUTIT A UNE LOI CANONIQUE

La possibilité même d'aboutir à une loi canonique peut pa-

raître surprenante. Elle signifie que si l'on applique la suite des opérations direct-inverse à un message d'information fixe, le produit des opérations ne ramène pas du tout au point de départ, la première fois qu'elles sont appliquées. Donc elles ne sont pas réellement inverses l'une de l'autre. (Mais si on les applique une deuxième fois, ou si on les applique dès le début à un message canonique, on revient au point de départ - donc le produit est une sorte de "projection").

Aucun phénomène de cette nature ne se produirait dans le codage optimum de Shannon. Du point de vue de ce codage, tous les messages sont équivalents, et il n'y a pas de correspondance inverse, car l'opération "inverse" est effectivement inverse de la directe, c'est-à-dire peut mener à n'importe lequel des messages initiaux. Avec ce codage, il n'y aurait donc pas de loi canonique.

Ainsi, les correspondances directe et inverse du codage mot par-mot sont toutes deux "singulières", chacune définissant une "équivalence" entre messages (resp. codage) conduisant au même codage (resp. message).

Toutefois, l'équivalence directe est dénuée d'intérêt, car elle réunit tous les messages où H et l'ordre des mots sont les mêmes, et ces messages sont très divers.

Au contraire l'équivalence inverse introduit le concept de "paramètre d'état" pour exprimer l'équivalence des messages.

Voyons comment ceci se présente si, à gain H donné, on représente graphiquement le coût en fonction des variables D = message, C = codage, qui sont les stratégies de deux interlocuteurs. Dans la zone C_S où C est une stratégie de Shannon, la surface représentative est un plan horizontal. Par contre, dans la zone C_M où C est une des stratégies mot-par-mot définies par diverses valeurs de E_g, toute coupe par un plan C = constant est une courbe située tout entière au-dessus du plan ci-dessus, et atteignant pratiquement pour la même valeur de D le même minimum de coût (mesuré en termes de - $\log_e$ M). Donc la surface possède une "vallée" rectiligne horizontale.

Le point optimum DC que nous cherchons est le point le plus bas de la surface de coût. Si les stratégies de Shannon sont permises, ce point est absolument indéterminé dans le plan C_S x D. Mais si c'est la zone C_M qui est seule permise, ce point n'est en pratique indéterminé que dans la vallée.

DECOMPOSITION NATURELLE DU COÛT CORRESPONDANT A CE CRITERE. En Mécanique Statistique, une fois que la distribution canonique a été obtenue, on constate que l'énergie se partage en énergie libre et en énergie liée.

De même ici, dans l'état canonique, le coût peut être décomposé en coût libre θH et coût lié. C - θH + Co. Dans ce contexte, le 1er critère économique a consisté à minimiser le coût lié à H donné. Ceci est tout à fait identique à ce qu'on fait en Thermodynamique, lorsqu'on minimise l'énergie libre. Dans les deux cas, on obtient un état stable. Mais les raisons du critère sont opposées. Là, on cherche l'état d'équilibre ou l'énergie utilisable est minimum, le système ayant déjà subi les pertes maxima compatibles avec les contraintes.

Ici, au contraire, la minimisation de la "perte", contre-partie de l'énergie utilisable, traduit précisément le fait que notre critère a pour but de donner l'état stationnaire de perte minimum.

Toute la différence entre les deux états stables provient de cette différence d'interprétation, due au décalage de toutes les quantités, coût remplaçant énergie, information remplaçant entropie et température thermodynamique étant remplacée par température informationnelle.

6.3.4 - ADAPTATION DE DEUXIEME TYPE

Elle peut se faire selon l'un des deux critères économiques ou selon un critère probabiliste. Avant de les poser, il faut considérer ce que serait la transmission à la Shannon sur q + 1 lignes. Alors M est remplacé par le nombre M', défini par $\sum_0^q M'^{-C_g} = 1$.

On a $r'_g = M'^{-C_g}$ $(0 \leqslant g \leqslant q)$ $\quad C'/H_e = (\log_e M')^{-1}$

Dans le <u>deuxième critère</u> économique, l'information par mot reste indéterminée. On devient alors libre de répartir le coût du top entre les R mots disponibles. Par exemple, on peut éviter d'avoir à utiliser les mots "compliqués" et rares en utilisant partout leurs définitions, ou inversement, etc... Pour cet émetteur prévoyant, la fréquence optimum minimum sera le coût moyen de l'unité d'information, les tops étant compris dans le coût.

On verra que ce critère fixe la quantité d'information par mot, qui ne sera pas la plus grande possible avec le nombre de mots dont on dispose, ce qui d'ailleurs n'est nullement en contradiction avec la théorie de Shannon.

Dans le <u>troisième critère économique</u>, l'information est indéterminée et on tient compte des tops, mais de façon différente. On minimise la différence entre :

- le coût moyen par mot, dans le codage mot-par-mot avec top, et
- le coût moyen dans le codage à la Shannon, sans tops. La ligne des tops est alors rendue libre de transmettre de l'information : en somme, le signal "ex-top" peut être utilisé au milieu des mots, et non plus seulement à la fin.

Les trois critères précédents sont économiques, tout comme le critère de 1er type.

On peut leur adjoindre un <u>critère probabiliste</u> (géométrique).

Représentons chaque pot par la suite de lettres donnée par le code du § 6.2.2. Même si les mots sont tirés au hasard, il n'y a aucune raison à priori pour que les lettres ou intervalles entre mots soient eux-mêmes tirés au hasard.

Or, chaque fois qu'il y a une liaison entre lettres, celles-ci apportent moins d'information qu'elles ne le pourraient s'il n'y avait pas de liaison (c'est d'ailleurs vrai aussi du point de vue naïf, car s'il y a une liaison, c'est que l'on sait déjà quelque chose à l'avance sur la lettre qui va venir).

Donc, pour qu'il y ait vraiment adaptation, il faut que la structure du message déjà codé soit aléatoire, intervalles compris.

6.3.5 - APPLICATION DES CRITERES DU DEUXIEME TYPE

Critère probabiliste.

Si le code qui représente le mot exige le top plus f symboles de coût total j_n, la probabilité du mot sera :

$$p_n = r'_0(1-r'_0)^{f-1}\,\Pi\,\frac{r_g}{1-r'_0} = \Pi\, r'_g\,\frac{r'_0}{1-r'_0}$$

Si l'on adopte les probabilités du § précédent,

$$p_n = \frac{r'_0}{1-r'_0}\,M'^{-j_n} \qquad (1 \leqslant n \leqslant \infty)$$

(On peut retrouver cette formule, en exigeant que la probabilité du mot n ne dépende que du coût de son codage, qui est la seule quantité rattachée à ce mot, qui ait un sens physique. De même, la probabilité de chaque symbole ne dépendrait que de son coût. Il en résulterait que r'_g devrait dépendre exponentiellement de C_g : $r'_g = M'^{-C_g}$, sans paramètre multiplicateur. La condition $\Sigma r'_g = 1$ redonne alors $\Sigma_0^q M'^{-C_g} = 1$: donc le M' doit être le même que dans le § précédent).

On peut écrire ce p_n sous la même forme que le p_n du § 6.3.3.

$$p_n = PM^{-Bj_n}$$

Mais, cette fois, ni R, ni H ne peuvent être choisis à priori : le critère détermine $R = \infty$ et $B = B_0 = (\log M')(\log M)^{-1} > 1$, donc H. En d'autres termes, on vient de trouver qu'il est impossible qu'un message doit aléatoire sur deux niveaux superposés distincts, sauf si sa structure au niveau supérieur est tout à fait particulière, dépendant d'ailleurs, non des détails de la structure inférieure, mais de fonctions de ces détails.

3ème critère économique.

Etant donné les coûts relatifs sur q et q + 1 lignes calculés au paragraphe précédent, l'excès de coût népérien dû à notre codage est :

$$(\log_e M)^{-1}\,\Sigma\, p_n\, j_n + (\log_e M')^{-1}\,\Sigma\, p_n \log_M p_n$$

Le minimiser redonne encore : $p_n = PM^{-B_0 j_n}$ avec le même B_0 que ci-dessus, mais R arbitraire et H et P fonction de ce R.

2ème critère économique.

Pour que $\dfrac{-\Sigma p_n \log p_n}{p_n\, j_n + C_0}$ soit minimum, à C_g donnés, il faut

$$(\Sigma p_n\, j_n + C_0)\,(-\Sigma \Delta p_n \log p_n + \Sigma \Delta p_n) + (\Sigma p_n \log p_n)\,(\Sigma \Delta p_n\, j_n)$$

$$-\,(1-K^2)\,(\Sigma p_n\, j_n + C_0)\,\Sigma \Delta p_n = 0$$

$$(\Sigma p_n\, j_n + C_0)\log p_n + K^2\,(\Sigma p_n\, j_n + C_0) + (\Sigma p_n \log p_n)\, j_n = 0$$

$$p_n = PM^{-Bj_n}$$

Le résultat est le même qu'au critère précédent, B est encore déterminé par l'ensemble des coûts, C_0 compris, mais par l'intermédiaire d'une équation caractéristique différente, à savoir :

$$B_e = \frac{H_M}{C + C_0} = \frac{-\Sigma p_n \log_M p_n}{p_n j_n + C_0} = -\frac{\log_M P}{C_0}$$

On peut résoudre en C_0 cette équation caractéristique :

$$C_0 = B_e^{-1} \log_M P = B_e^{-1} \log_M \Sigma_1^R M^{-Bj_n}$$

Supposons que l'on ait le codage simple. Alors :

$$C_0 = B_e^{-1} \log_M \Sigma_1^J M^{-(B_e-1)J} = B_e^{-1} \log_M \frac{M^{(1-B_e)}(1-M^{(1-B_e)J})}{1-M^{(1-B_e)}}$$

$$= V\ (B_e,\ M,\ R) \quad \text{où} \quad J = j_R$$

V est une fonction croissanre de J et R et une fonction décroissanre de B_e, c'est-à-dire que plus le coût du top est grand, plus grande doit être l'information, c'est-à-dire plus il faut attendre en moyenne avant d'utiliser ce top.

V est infinie pour $B_e = 0$, passe par $\log_M \log_M R$ pour $B_e = 1$ et descend à 0 pour B_e, tel que $M^{(1-B_e)} = D$ soit donné par l'équation $D^{J+1} - 2D + 1 = 0$. Pour $J = \infty$, $D = 1/2$, ce $B_e = 1 + \frac{1}{\log_2 M}$.

<u>Coût népérien redondant</u> : C'est $C_0 \log_e M + \log_e P$

<u>Redondance du codage</u> : C'est $(C_0 + \log_{M'} P)\ (C + C_0)^{-1}$

<u>Rendement du codage</u> : C'est $H_{M'}\ (C + C_0)^{-1} = \frac{-\log_{M'} P + C}{C_0 + C}$

Dans le cas du 2ème critère, ceci se réduit à B_e/B_0. Vérifions que ce nombre est bien < 1, à C_0 donné, c'est-à-dire que pour B donné, l'on a :

$$\frac{1 - M^{(1-B)}}{M^{(1-B)}\ (1 - M^{(1-B)J})} > 1 - \Sigma_1^q M^{-BC_g}$$

Le terme de gauche étant fonction décroissante de J, il suffit de démontrer l'inégalité pour $J = \infty$. Or pour $B = 1 + \varepsilon$, les deux termes donnent tous deux $\varepsilon \log M$: ils sont égaux. Autrement dit, pour C_0 très grand, la réaction de C_0 sur H ne dépend pas du critère. Pour B croissant, le 1er terme est concave vers le haut, le 2ème vers le bas, donc l'inégalité est conservée, CQFD.

<u>Rendement apparent du top</u>. Nous appellerons ainsi le rapport $-\log_M P/C_0$. Ce rapport est négatif pour B suffisamment grand, mais il est positif pour B plus petit, par exemple pour $B = 1$. Alors $C < H$ et les tops transmettent de l'information, au lieu d'être uniquement symboles de synchronisation. Ce sont des symboles exactement aussi importants que les autres (la possibilité de comprendre un texte réel, écrit sans intervalles entre mots n'a pas plus de signification que la possibilité de supprimer tout autre signe dans un texte redondant).

<u>Rendement réel du top</u>. Pour l'estimer, il faut tenir compte de la diminution d'équivocation, à probabilité d'erreur sur les

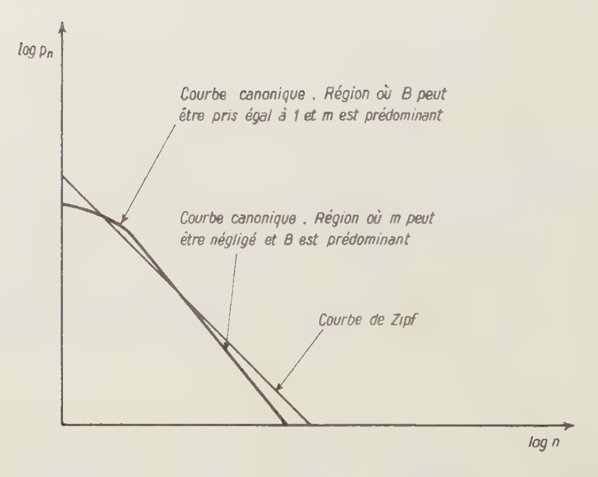


Fig. 1. Distribution canonique sans fréquence $p_n = P(n + m)^{-B}$
Comparée avec la distribution de Zipf $p_n = Pn^{-1}$

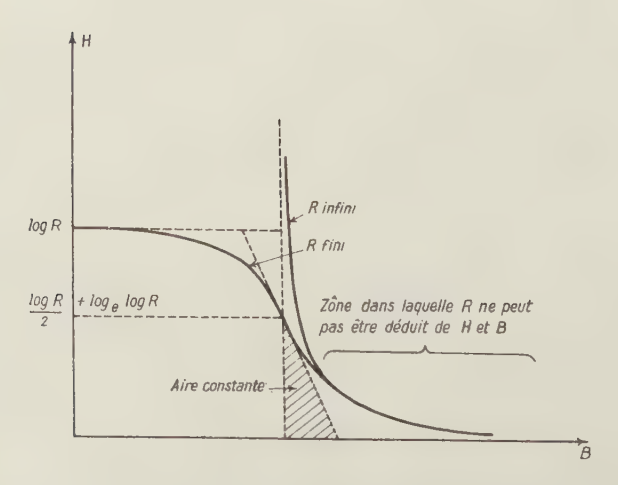


Fig 2. Fonction d'état H. Sa variation
en fonction des variables d'état B etR

symboles élémentaires donnée, lorsque l'on passe du codage à la Shannon au codage mot par mot. Alors $C + C_o$ doit être diminué de l'équivocation mot par mot, et $- \log P$, de l'équivocation à la Shannon et le rendement augmente.

Rôle correcteur du top.

Le top a d'abord été introduit pour éviter qu'une seule erreur de transmission (ailleurs que sur le top) ne détruise tout le message. A priori, il n'y avait aucune raison pour que l'on ait la possibilité de corriger cette erreur. Mais en fait, on a le résultat plus fort suivant : Pour les messages satisfaisant à la loi canonique, il existe un code correcteur d'erreur simple, qui exige - d'une part un nombre de symboles sur des lignes "information" strictement égal à l'information transmise, - d'autre part un seul symbole additionnel sur une ligne additionnelle, servant simultanément de top de synchronisation et de symbole correcteur d'erreur unique. (Posons le problème de savoir si des codes correcteurs de n erreurs existent qui exigent n lignes additionnelles).

Ce code résulte d'une modification du code de R.W. Hamming (1951) qui tient compte de la possibilité d'envoyer le top à tout instant et pas seulement à intervalles réguliers.

La modification résulte du double rôle du "zéro". Dans Hamming, tous les nombres sont complétés par des zéros pour avoir des longueurs égales : les zéros ne transmettent pas d'information, mais peuvent être retransmis comme "Un"'s, donc transportent de l'erreur potentielle. Donc, dès que la nécessité de les envoyer est supprimée par les tops mobiles, le nombre de symboles à transmettre diminue beaucoup.

Il faut toujours n symboles pour indiquer les erreurs simples dans $2^n - 1$ positions. Ici on peut utiliser 1 symbole correcteur pour les mots de 1 symbole, 2 symboles pour les mots de $2^2 - 1 - (2 - 1) = 2$, 3 symboles pour les 5 suivants, etc...

Si les fréquences des mots suivent la loi canonique pour $B = 1$, c'est-à-dire $p_n = 1/n \log_e R$, le nombre de symboles d'information est $(\log R / 2$ et le nombre de symboles correcteurs $\log \log_e R$ la somme est bien égale à l'information :

$$\log R / 2 + \log \log_e R \quad CQFD$$

La condition : erreur unique se traduit difficilement en équivocation, ce qui rend difficile la définition du rendement de ce codage.

6.3.6 - AUTRES CRITERES

Bien d'autres critères, plus ou moins justifiables, font retomber sur la valeur $B = 1$ qui est celle autour de laquelle se répartissent d'habitude les valeurs des critères précédents.

N. Rashewsky (1950) obtient $H = \log_2 \sqrt{R}$, en maximant $(\log R - H) H$, produit de la quantité d'information par la "quantité d'ignorance" lorsque l'on connaît R. Etant donné qu'il ne considère pas de statistique particulière p_n, le point représentatif du message dans l'espace dont les R axes portent les probabilités reste sur une hypersurface à $R - 1$ dimensions; mais parmi les messages de cette surface, le seul qui soit canonique est $B = 1$.

6.4 - LA LOI CANONIQUE

6.4.1 - PROPRIETES DE LA LOI CANONIQUE

Les résultats du chapitre précédent montrent que quel que soit le critère d'adaptation, il conduit à une loi appartenant à une famille unique :

$$p_n = P \quad [n + m]_M^{-B}$$

où $[x]_M$ est tel que $\log_M [x]_M = [\log_M x + j_0]$

En fait, il ne s'agira pas en pratique de construire un message adapté au codage arithmétique, mais il faudra reconnaître à posteriori, d'après un échantillon, si un message réel est ainsi adapté, lorsqu'on le divise en "mots" d'une certaine façon, connue à priori. Or s'il l'était, la dispersion et le reclassement par probabilités décroissantes amortiraient les marches d'escalier de la courbe p_n ci-dessus, laissant une distribution de fréquences très proche de la "loi canonique simplifiée" :

$$\boxed{p_n = P\,(n + m)^{-B}} \quad \text{(fig. 1, p. 94)}$$

où : $1 \leqslant n \leqslant R$, et $P^{-1} = \Sigma_1^R (n + m)^{-B}$

Les effets de M sont secondaires et seront étudiés séparément (§ 7.4).

En coordonnées bilogarithmiques :

$$\boxed{-\log p_n = -\log P + B \log (n + m)}$$

C'est une droite de pente -B, sauf pour les petites valeurs de n, où les mots sont moins fréquents que ne l'aurait fait croire cette droite. On peut considérer deux approximations distinctes à la loi canonique :

A) <u>pour n grand</u> : m est négligeable et ce qui est prépondérant c'est B, c'est-à-dire le message, et plus précisément la longueur finie des tranches d'information et le pourcentage du coût consacré (inévitablement) aux tops. On écrira $\boxed{p_n = Pn^{-B}}$.

On a une divergence par rapport à la loi de Zipf $P_n = Pn^{-1}$.

B) <u>Pour n petit</u> : si B est suffisamment proche de 1, l'on peut négliger son influence, c'est-à-dire celle du message, et l'assimiler à 1, précisément dans la zone où m doit être conservé. Alors, intervient surtout l'influence de m, c'est-à-dire du système de codage; si on pouvait mesurer m on aurait des indications sur les détails de ce système. On écrira $\boxed{p_n = P\,(n + m)^{-1}}$.

m provoque une divergence par défaut par rapport à la loi de Zipf : $p_n = P'n^{-1}$ de même sens et de même nature que la divergence de la loi de Fermi-Dirac par rapport à la loi de Boltzmann.

De tels coefficients m se retrouvent très fréquemment, en physique, physiologie et psychologie. Tous les exemples font intervenir directement, ou dans des modèles physiques, une quantification du type de celle qu'effectue le codage arithmétique. Il est donc tentant de voir dans celui-ci l'origine de m. Il est d'ailleurs évident que dès que l'on fait tendre tous les "quanta" vers 0 sans changer l'ordre de grandeur des quantités physiques observées, la mesure de celles-ci au moyen de ces quanta augmente indéfiniment et à côté de cette mesure m devient absolument négligeable.

L'exemple fondamental de m se rencontrera dans l'étude statistique de la langue au Chapitre 7 : la loi canonique est satisfaite, sans que le coût puisse être mesuré. Mais nous avons déjà au § 6.2.2 cité un exemple auxiliaire, dû à Hick, où le coût est mesuré directement. Par ailleurs m semble apparaître dans certaines "améliorations" de la loi de Weber et Fechner. Mais la "sensation" n'étant pas quelque chose de mesurable, nous laisserons cet exemple de côté.

6.4.2 - VARIATION EXACTE DE H EN FONCTION DE B (FIG. 2 P. 94)

$$H = \Sigma_1^J PM^{-Bj} \log (PM^{-Bj}) M(j) \quad \text{avec} \quad J = \log_M R.$$

$$= - \log P + PB\, \Sigma_1^J j \log M\, P\, M^{-Bj} M(j)$$

$$= - \log P + B \frac{d}{dB} \log P$$

Dans le cas simple :

$$P^{-1} = \Sigma_1^J \dot{M}^{-Bj} M(j) = \Sigma_1^J M^{(1-B)j} = M^{(1-B)} \frac{M^{(1-B)J} - 1}{M^{(1-B)} - 1}$$

$$H = (1 - B) \log M + \log \quad M^{(1-B)J} - 1 \quad - \log \quad M^{(1-B)} - 1$$

$$+ B \log M + B \frac{J \log M\, M^{(1-B)J}}{M^{(1-B)J} - 1} - B \frac{M^{(1-B)} \log M}{M^{(1-B)} - 1}$$

Dans tous les cas, la courbe H (B) est évidemment continue :

$$\frac{dH}{dB} = (\log^2 M)\, B \left\{ (\Sigma p_n f_n)^2 - \Sigma j_n^2 p_n \right\}$$

est $\leqslant 0$ par l'inégalité de Schwartz, donc H décroît continument de log R à 0 quand B va de 0 à ∞.

Près de B = 1, la variation est très rapide $H = \log \log R + \frac{\log R}{2}$

$$\frac{dH}{dB} = \frac{-(J - 2)^2}{12} + \frac{7}{12}$$

Si $B > 1$, lorsque $J \to \infty$, $M^{(1-B)J} \to 0$;

$$\log (1 - M^{(1-B)J}) \quad - M^{(1-B)J}$$

$$H = - \log M - \log (1 - M^{1-B}) - B \frac{\log M\, M^{(1-B)}}{1 - M^{(1-B)}}$$

$$- M^{(1-B)J} + BJ \log M\, M^{(1-B)J}$$

H augmente avec J et tend vers une limite finie indépendante de R :

$$H = -\log M - \log(1-M^{(1-B)}) - B\frac{\log M\, M^{(1-B)}}{1-M^{(1-B)}}$$

Si $B - 1 \ll 1$, et $B > 1$;

$$H = -\log M - \log(B-1)\log M - \frac{B}{B-1} + \log M$$

$$= -\log M - \log\log M + B\log M - \frac{B}{B-1} - \log(B-1)$$

Pour $B < 1$, $M^{(1-B)J}$ 1, lorsque J est grand (mais fini)

$$H = -\log M + (1-B)J\log M - \log(M^{(1-B)} - 1$$

$$+ BJ\log M + B\frac{M^{(1-B)}\log M}{M^{(1-B)}-1}$$

$$= \underbrace{-\log M}_{\text{constante absolue}} + \underbrace{J\log M}_{=\log R} - \underbrace{\log(M^{(1-B)}-1) + B\frac{M^{(1-B)}\log M}{M^{(1-B)}-1}}_{\text{constante relative}}$$

constante absolue — variable avec R : indép. de B — constante relative

Si $B - 1 \ll 1$, et $B < 1$;

$$H = -\log M + \log\log M - \log(1-B) + \frac{B}{B-B} + B\log M + \log R$$

Pour $J \gg 1$; la pente pour B - 1 est $\frac{J^2}{1^2}$ et l'ordonnée $\frac{J}{2}$. La tangente intercepte alors $\frac{6}{J}$ sur les ordonnées 0 ou J; et l'aire comprise entre la tangente, ces ordonnées, et l'abscisse B = 1 est constante.

De même C est une courbe continue variant de log R à 0 en passant par $\frac{\log R}{2}$ pour B = 1, valeur près de laquelle elle varie très vite.

Sa dérivée est égale à celle de H, divisée par B.

$$\frac{d^2C}{dB^2} = \frac{d^3}{dB^3}\log P;\quad \frac{d^2H}{dB^2} = \frac{d^3}{dB^3}\log P + \frac{d^2}{dB^2}\log P$$

$$\text{or}\quad \frac{d^2}{dB^2}\log P = \frac{J^2\log^2 M}{(1-M^{(B-1)J})^2} - \frac{\log^2 M}{(1-M^{(B-1)})^2}$$

$$\text{et}\quad \frac{d^3}{dB^2}\log P = \frac{2J^3\log^3 M}{(1-M^{(B-1)J})^3} - \frac{2\log^3 M}{(1-M^{(B-1)J})^3}$$

sont tous les deux nuls pour B suffisamment proche de 1 (mais ≠ 1 dans cette formule), ainsi que toutes les dérivées d'ordre supérieur. Ainsi pour B = 1 H (B) présente un point d'inflexion d'ordre.

Le contact est donc très étroit et la courbe se confond avec sa tangente sur une grande étendue de B.

6.4.3 - % D'UTILISATION DU VOCABULAIRE ET ECONOMIE PAR LA METHODE SEQUENTIELLE

Nous appellerons pourcentage d'utilisation du vocabulaire l'expression $H_e/\log_e R$. C'est une fonction de B et R. Pour B = 1

$$\frac{\frac{\log R}{2} - \log \log_e R}{\log R} = \frac{1}{2} = \frac{\log \log_e R}{\log R}$$

donc finalement de l'ordre de $\frac{1}{2}$.

Cette notion est conceptuellement différente de celle de redondance : son inverse donne l'expansion que provoquerait le remplacement du codage séquentiel par un codage non séquentiel, où tout mot aurait la longueur log R. Cette expansion serait donc du double pour B = 1, ce qui est considérable.

Elle se trouve identique au résultat moyen que Wald (1947) annonce avoir trouvé en statistique proprement dite, là où il est intéressant d'utiliser l'analyse séquentielle. Cette "coïncidence" peut s'expliquer de la façon suivante : dans les cas où l'on peut supposer que les combinaisons de résultats d'expériences séquentielles susceptibles d'être obtenues sont en grand nombre R :

Le nombre R n'est facile à déterminer que si le log de la probabilité de l'alternative le moins probable est de l'ordre de grandeur de la quantité d'information. C'est par exemple le cas si un B est défini et inférieur à 1 : dans ce cas l'économie par analyse séquentielle aurait été très faible, mais personne ne songerait à utiliser l'analyse séquentielle.

Si au contraire $-\log p_R \gg H$, une forme d'analyse séquentielle s'impose d'elle-même. En effet, R étant mal déterminé, toute analyse non séquentielle basée sur une valeur définie de R serait de toute façon insuffisante, sauf si on prend R très grand, auquel cas l'analyse aurait un rendement très mauvais.

Par suite, les seuls cas où il y a choix entre méthode séquentielle et méthode non séquentielle, donc où la comparaison ait été faite entre ces méthodes, sont ceux où on se trouve entre les deux cas précédents. Or si la série est juste convergente ou juste divergente, les sommes $\Sigma p_n \log n$, $\Sigma p_n \log p_n$ sont peu sensibles à la forme de p_n tant que l'on a des séries marginales. Il est par suite plausible que, si l'économie est toujours de l'ordre de $\frac{1}{2}$, c'est que l'on ne considère que les cas où ce phénomène est susceptible de se manifester. Ce $\frac{1}{2}$ est donc souvent explicable sans qu'il y ait besoin de faire une théorie très précise de chaque cas particulier.

6. 5 - REMARQUE : PROCESSUS MODÉRÉS

6.5.1 - CRITERES D'ENTROPIE MINIMUM

Revenons maintenant à l'interprétation entropique du coût. Nous avions rematqué au § 0.3.4 (fin) que les signaux qui interviennent dans la reconnaissance de la langue sont très supérieurs au minimum kT. Il n'en demeure pas moins que le critère de base du § 6.3 est un critère de minimum appliqué à l'augmen-

tation d'entropie du décodeur consécutive au décodage. Nous avons longuement montré que le principe de meilleure économie détermine l'augmentation d'entropie dans un cas où le principe de Carnot ne disait rien, sinon qu'elle était positive.

Le principe de Carnot n'est en effet qu'un principe de discrimination entre processus impossibles, réversibles et naturels. Il ne détermine donc entièrement un processus que lorsque la variation d'entropie ΔS correspondante est déterminée par des conditions extérieures, seule la direction restant indéterminée. Tel est en particulier le cas des ondes de choc en Mécanique des Fluides.

Il existe cependant de nombreux cas où l'augmentation d'entropie est en principe indéterminée à une inégalité près, mais où l'expérience prouve qu'en réalité les processus sont déterminés, et même stables (revenant à l'état initial s'il a été dérangé) et modérés (répondant à un changement de paramètres extérieurs d'une façon qui, agissant seule, aurait provoqué le changement inverse de ces paramètres : principe de Le Châtelier Braun).

Faut-il considérer ces processus comme des cas pathologiques ou peut-on les déterminer par un ou plusieurs critères tout à fait généraux. Il est impossible de répondre : on ne peut que signaler que des processus extrêmement variés peuvent être déterminés par le plus simple des critères concevable : - la condition que l'augmentation d'entropie soit minimum, étant données les contraintes et les principes de bilan appropriées au processus étudié.

C'est aussi une condition de moins mauvaise utilisation de la durée.

On pourrait peut-être conjecturer que les seuls processus irréversibles, susceptibles d'étude quantitative, sont ceux qui satisfont à ce critère.

Prigogine et Wiame (1948) ont d'ailleurs proposé d'étudier ainsi des processus vivants, sans d'ailleurs donner d'exemple quantitatif d'une telle étude. Pour eux, "le caractère adaptatif très général des organismes qui se manifeste dans la forme, autant que dans le mécanisme physiologique, peut être considéré comme une tendance de la matière vivante à effectuer un travail maximum avec une dépense de matière minimum".

Notre théorie de la langue du Chapitre 7 a été réalisée sans connaître leur travail; mais elle montrerait la justesse de leur conjecture dans un cas particulièrement typique. Par ailleurs, bien que le critère de base n'ait été inspiré par aucun des autres processus irréversibles, le rôle qu'il jouera ne sera pas à posteriori sans analogies.

6.5.2 - EXEMPLES

Exemple de la détonation de Chapman-Jouguet (Cf. Courant et Friedrichs 1948).

Le processus de combustion d'un gaz dans un tube n'est pas déterminé par les lois de l'énergétique plus la condition $\Delta S > 0$ (comme l'est par exemple une onde de choc). La combustion progressive et lente est métastable, et peut se transformer en dé-

tonation très rapide, déterminée par l'une de plusieurs conditions équivalentes, dont la loi de Jouguet : La déflagration correspond à l'augmentation minimum d'entropie par unité de masse.

Exemple des processus étudiés par Onsager (cf. Casimir) (1946), Prigogine (1947) et de Groot (1951).

Ces systèmes évoluent en général vers des états stationnaires, qui correspondent à une production minimum d'entropie, compatible avec les conditions imposées au système, et qui sont stables. Ex. Effet Knudsen : 2 compartiments de gaz sont à la même piession et à des températures différentes. Un trou est percé dans la paroi qui les sépare. Les pressions deviennent différentes de façon à rendre minime la production d'entropie par unité de masse due aux transports de matière et de chaleur.

6.5.3 - AUTRES ANALOGIES ENTRE CES PROCESSUS

Dans le cas du couple $(D. D^{-1})$ optimal du Chapitre 5, on n'a d'abord aucun changement d'entropie, puis lorsque le message est démoli, brouillé, etc... on a une augmentation égalr à l'information perdue. Donc l'irréversibilité se fait au cours d'une deuxième étape du processus, sans influence sur la première.

Il est tout à fait légitime de se demander si la possibilité d'étudier les autres processus ne vient pas aussi de ce qu'ils se font en deux étapes, dont la deuxième, seule irréversible, n'influe pas sur la première, qui est une création d'ordre mesuré par le minimum d'entropie.

Effectivement, la détonation de Chapman-Jouguet correspond entre autres propriétés à une vitesse sonique de la discontinuité par rapport aux gaz brûlés, ce qui explique la non-influence en amont des irréversibilités dues aux conditions finales.

Par ailleurs, dans les processus étudiés par Onsager, on utilise la formule locale de Gibbs, ce qui revient implicitement à considérer les différences entre les grandeurs locales (mais non les gradients) et par suite à tenir compte de l'ordre dynamique stationnaire. (En théorie de l'information, on ne tient pas non plus compte des gradients).

EXEMPLES DE DONNEES NUMERIQUES TIRES DE ZIPF (1949)

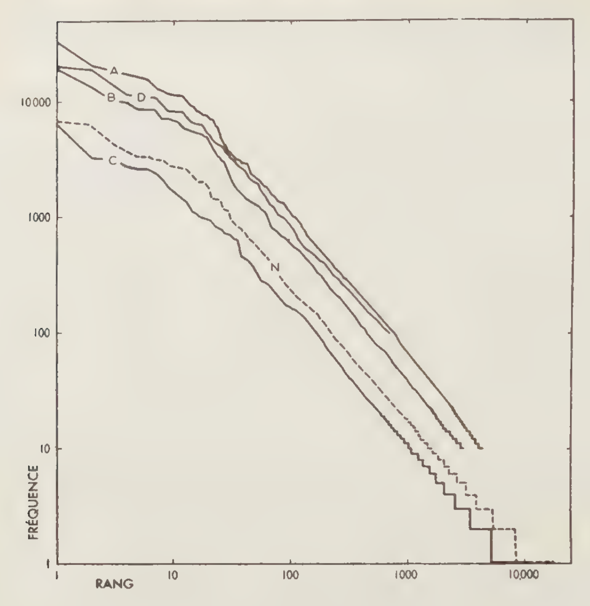


Fig 3a. A-C Norvégien
N : Allemand Notker (cf Fig 3c)

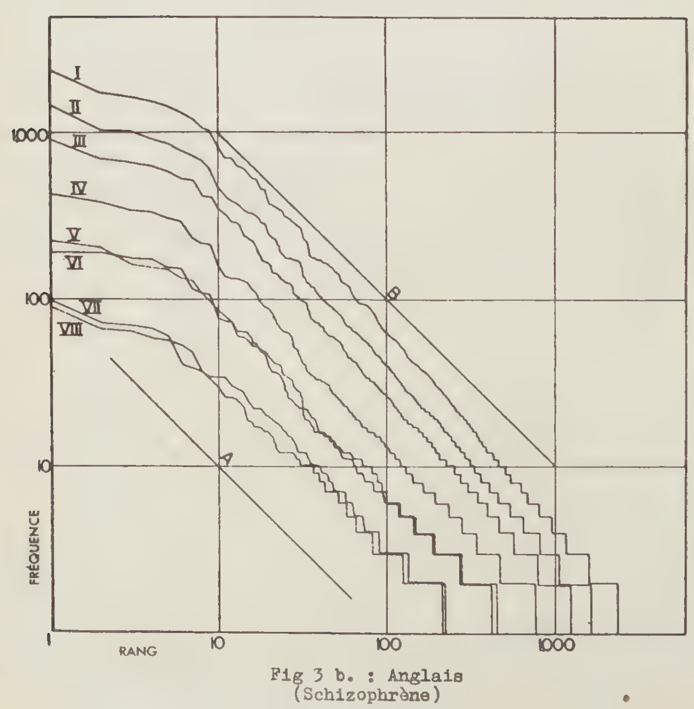


Fig 3 b. : Anglais
(Schizophrène)

CHAPITRE 7

STRUCTURE STATISTIQUE DE LA LANGUE

7. 1 - LOI CANONIQUE EN LINGUISTIQUE

7.1.1 - RESULTAT FONDAMENTAL

Rappelons ici le résultat fondamental du § 0.3.4, à savoir : que la langue, considérée comme suite de mots entièrement fléchis, possède une structure statistique canonique.

Nous allons tirer les conséquences de ce fait.

La loi canonique a pour premier résultat de permettre de partager la description de la statistique en trois éléments :

- d'une part, l'ordre des mots par fréquence décroissante,
- d'autre part, quelques paramètres du codage,
- enfin, quelques paramètres du message.

Il faut que ces trois éléments soient les mêmes pour l'émetteur et le récepteur pour qu'il y ait entre eux adaptation, c'est-à-dire coalition.

La possibilité de distinguer entre les 2° et les 3° éléments traduit le fait très important que la présente coalition peut se décomposer en deux éléments:

- l'accord nécessaire, et bien connu, entre encodage et décodage
- la propriété de minimin, dégagée dans ce travail.

7.1.2 - LE POINT DE VUE MACROSCOPIQUE

Les méthodes qui conduisent à la loi canonique étaient microscopiques, c'est-à-dire qu'elles considéraient le texte comme une suite de mots. Cependant, pour de nombreux usages, et en particulier le calcul de tous les nombres rattachés au texte, seule est nécessaire la troisième partie de la description ci-dessus du vocabulaire.

La possibilité de dégager une description moins spécifique, mais suffisante, macroscopique, constitue un "trait fascinant" que la théorie de l'information partage avec la thermodynamique.

EXEMPLES DE DONNEES NUMERIQUES TIRES DE ZIPF (1949) (suite)

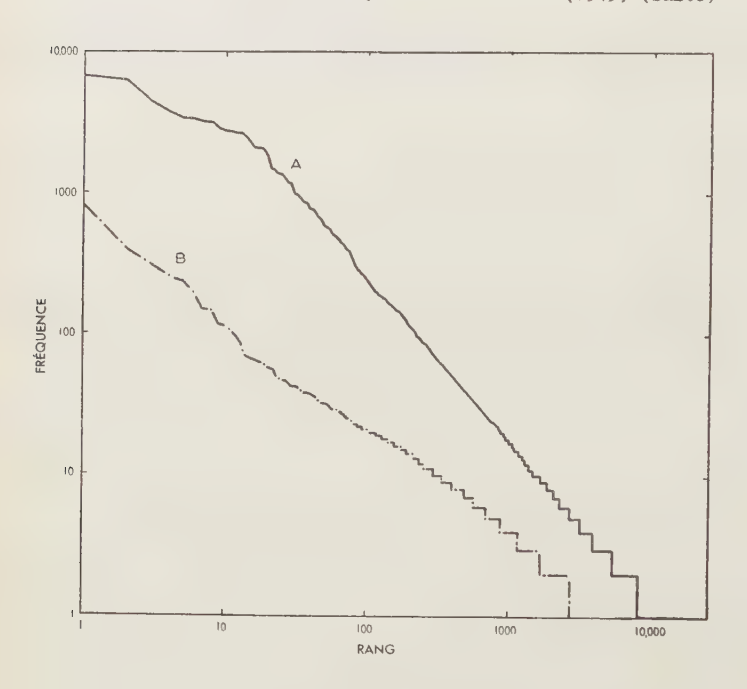


Fig 3c : Notker
A : allemand B : latin

EXEMPLES DE DONNEES NUMERIQUES TIRES DE ZIPF (1949) (suite)

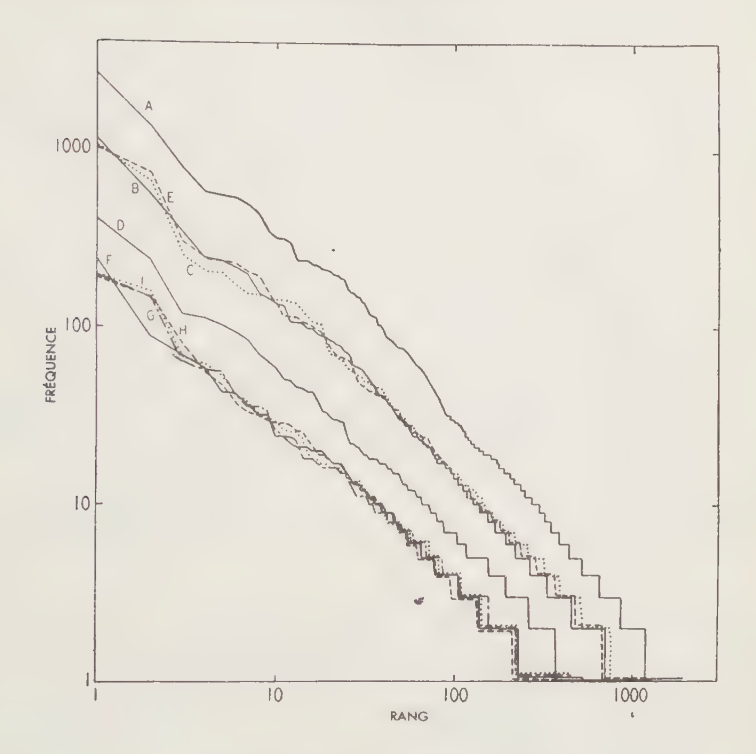


Fig 3d. Enfant anglais

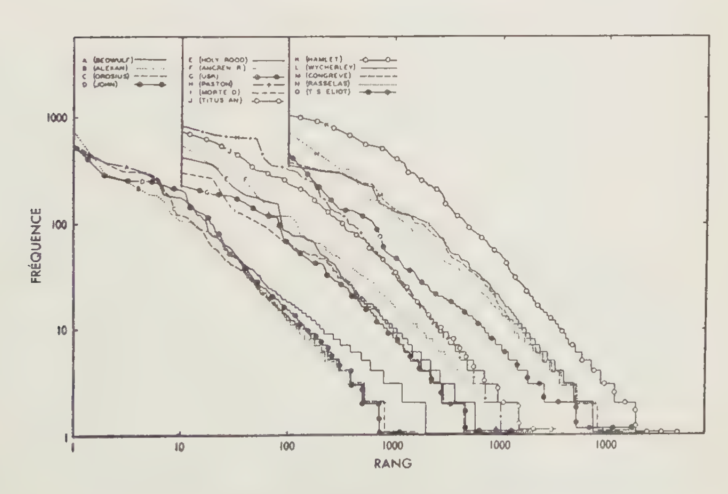


Fig 3e : Beowulf à T.S Eliot
Quinze auteurs anglais

Toutes les deux, selon les termes de E. Schödinger (1948) "donnent presque invariablement un sens physique et fondamental à des quantités purement mathématiques par leur définition".

Une description macroscopique présente également l'avantage de permettre d'estimer les répercussions que produiraient des modifications de la localisation et de la direction des paramètres libres du système, sans avoir besoin de se référer à chaque fois au substratum exact. Cette description rend donc possible l'acquisition d'une "intuition" de la structure du texte.

La description macroccopique présente enfin un avantage esthétique : celui de donner à la théorie d'un message canonique une unité et une personnalité propres, qui se rattacheraient difficilement à une liste de p_n.

Dans le chapitre précédent (§ 6.3.2), nous avons déjà vu un début d'exploitation de l'analogie macroscopique entre la théorie des messages canoniques et la thermodynamique. Nous continuerons ici, en introduisant des dénominations et des dimensions "informationnelles" macroscopiques inspirées de la thermodynamique, pour désigner les nombres purs, caractérisant la structure élémentaire, microscopique du message (variable d'état), et les expressions construites à partir de ces nombres (fonctions d'état). Il est probable que ces dénominations pourront éclairer notre problème et suggérer par analogie de nouveaux développements.

L'analogie avec la Thermodynamique ne s'étendra pourtant pas aux techniques de mesure : dans ce domaine, une opposition subsistera en pratique entre Thermodynamique et Théorie du message canonique, par suite de la différence entre les dimensions de l'objet par rapport aux instruments possibles (mais cette opposition n'a pas de conséquences théoriques graves). En effet, le thermomètre (qui est l'instrument type de la Thermodynamique) est à une échelle beaucoup plus grande que les molécules dont il mesure l'énergie moyenne pour obtenir la température. Par contre, la mesure des fonctions et des variables d'état informationnelles se fait habituellement en effectuant implicitement la statistique du texte et tabulant ensuite des fonctions des probabilités. Donc la moyenne ne se fait pas automatiquement, ce qui équivaut à compenser les effets de fluctuations en "promenant" un instrument trop petit (comme on n'aurait à le faire en thermodynamique que s'il était concevable qu'un thermomètre soit infiniment petit).

Toutefois, dans la mesure où les textes ne diffèrent pas par l'ordre des mots par fréquence décroissante, mais seulement par B, il existe un répertoire de mots universel. On peut alors imaginer un instrument qui remplacerait au fur et à mesure chaque mot par son code optimum. La moyenne des coûts se ferait automatiquement sans tabulation et on en déduirait par exemple B, par une graduation adéquate de l'échelle des coûts. Il n'est pas absurde de croire que c'est ainsi que l'on perçoit la "difficulté d'on vocabulaire" à l'aide d'un "instrument" cérébral.

7.1.3 - SEMANTIQUE

La description macroscopique comporte une simplification extrême de la description de la langue, qui consiste à pousser jusqu'à la limite la théorie structurale de F. de Saussure

(1916). Une telle simplification ne doit nullement être considérée comme une fin en soi. Elle est cependant nécessaire dans cet essai pour utiliser en linguistique des notions mathématiques.

Une simplification différente du langage, tout aussi extrême, quoique en sens inverse que celle qui conduit à la langue selon de Saussure, est celle qui a conduit à la Sémantique de l'Ecole de Vienne (Schlick, Carnap, Reichenbach). Pour ceux-ci, la science des signes a trois subsidivions :

- Pragmatique (référence explicite à un émetteur)
- Sémantique (référence aux choses désignées, mais non à l'émetteur)
- Syntaxe (référence aux expressions seules).

La linguistique contiendrait la pragmatique et les parties descriptives de la sémantique et syntaxe.

La sémantique et syntaxe pures sont allées si loin dans l'abstraction, qu'elles n'ont laissé que peu de possibilités de prévoir des faits observables, tels que ceux auxquels nous a conduit notre théorie.

Il aurait été souhaitable de joindre les deux cas extrêmes, par exemple en cherchant les probabilités de passage. Mais dès qu'on essaye d'introduire la sémantique par morceaux, elle fait irruption tout entière.

7.1.4 - REDONDANCES

Au § 6.3.5, nous avons calculé la redondance du codage optimal mot par mot. Il faut éviter de la confondre avec la redondance qu'implique l'expression des idées par les mots. Une telle nouvelle notion serait de caractère sémantique. Si on introduit cet aspect, les redondances se multiplient. On obtient même des notions distinctes selon qu'il s'agit de la source ou de l'observateur. A la réception, il faut tenir compte du "message" : "non compris". La probabilité de ce signal est aussi une redondance !

Le "pourcentage d'utilisation" des mots H/log R pour R est fini, peut-être aussi confondu avec 1 - redondance. Si $R \to \infty$, et $B > 1$ ce % $\to 0$, mais on peut comparer les % de deux textes en prenant le rapport de leurs informations.

Cette multiplicité de définitions rend très incertaines les mesures expérimentales de la "redondance intuitive", car on sait alors très mal ce que l'on mesure : un message non redondant à un point de vue peut l'être à un autre.

Par exemple, Shannon (1951) déclare que d'après les processus de suppression et de reconstitution, un texte de James Joyce est beaucoup moins redondant qu'un texte en Basic English : cela pourrait être dû à l'inégale expansion qu'introduit l'orthographe pour les mots courts et les mots longs, mais plus vraisemblablement au fait que l'observateur commun utilisé pour comparer les deux textes a des p_n nettement différents de l'un comme l'autre texte, donc qu'il perçoit aussi le % d'utilisation des mots. Pour un observateur hypothétique accordé au texte, la redondance de celui-ci dans le dernier sens serait beaucoup plus faible.

7.2 - PARAMÈTRES MACROSCOPIQUES DU MESSAGE

7.2.1.

La plus importante des fonctions d'état du message est la "quantité d'information sélective" H définie par la formule = $H = - \sum p_n \log p_n$ qui est la contrepartie exacte de l'"entropie" de la Thermodynamique.

Cette espèce d'information est donc parfaitement définie et ne conserve rien de l'imprécision habituelle du terme "information". Elle ne représente d'ailleurs qu'une petite partie du sens de cette "information", à savoir l'élément de surprise qu'apporte la réception d'un mot.

Parmi les fonctions d'état, citons aussi la "somme d'état" P (constante de la loi canonique) et le coût moyen du mot en codage optimum.

Quant aux variables d'état du message à étudier elles sont au nombre de deux : B, R.

Le paramètre B se mesure aisément comme pente de la courbe $- \log p$ en fonction de $\log n$ pour n grand. Par contre, le nombre de mots potentiel R n'est pas donné, ni mesurable directement, sauf dans le cas complètement irréel où l'échantillon étudié contient chaque mot au moins une fois, c'est-à-dire où le nombre de mots différents cesse d'augmenter à partir d'un certain point. Il faudrait disposer d'un texte infini pour pouvoir affirmer ce fait avec certitude.

Si on ne dispose pas d'un tel texte, la valeur de R est matière à estimation à partir des fonctions d'état mesurables et il est nécessaire de préciser l'importance que des erreurs d'estimation sur R auront sur les valeurs des fonctions mesurables dont le calcul exige R. L'importance des erreurs sur R, donc l'importance de R sera très grande ou insignifiante selon que B sera inférieur ou supérieur à sa "valeur critique" B = 1.

Comme en fait $B > 1$ dans presque tous les cas, c'est B qui est la plus importante des variables d'état.

7.2.2 - INFLUENCE DE B SUR L'ESTIMATION DE R DANS LE CAS OU B > 1

(Ce vocabulaire sera appelé ouvert, pour des raisons qui deviendront évidentes par la suite). Dans le cas où $B > 1$, on n'a aucun besoin de s'assurer effectivement d'une valeur déterminée de R. En effet, toutes les fonctions d'état sont données par des sommes finies qui peuvent être considérées comme des sommes partielles de séries très rapidement convergentes. (Tel est toujours le cas en Mécanique statistique des Gaz). Par suite, ces sommes peuvent être assimilées aux sommes infinies de ces séries, même pour R modérément grand.

Par suite, R n'influe pas sensiblement sur le coefficient P, ni sur la quantité d'information, ni sur la longueur moyenne des mots dans le codage optimum. L'on pourra changer l'estimation provisoire de R, sans changer ces sommes.

Le seul cas où l'on aurait pu craindre des difficultés est celui où le R est estimé petit et B petit aussi. Alors on n'au-

rait pas pu légitimement remplacer les sommes finies par les sommes des séries. Fort opportunément, il "se trouve" que ce cas ne se rencontre pas, le "vocabulaire pauvre" coïncidant toujours avec "faible information par mot", donc B grand, cas auquel la valeur de R peut être petite sans inconvénient. Il est à cet égard intéressant de constater que l'accord reste bon dans deux cas opposés où la quantité d'information est faible : dans un texte populaire, où le nombre de mots différents semble petit, et chez les schizophrènes, où le nombre paraît plus grand.

Si, le texte croissant, on doit introduire de nouveaux mots, on n'aura pas à remettre en question les fréquences des mots anciens; le texte s'allongeant, les fréquences tendent vers des limites déterminées et la notion de probabilité peut prendre tout son sens dans un seul texte. R n'influant pas sur le calcul des quantités relatives au vocabulaire, il n'y aura pas inconvénient, en général, à supposer infinie cette quantité "indifférente" ou "indéterminée".

Toutefois, il existe une limite supérieure à R constituée par le nombre R_0 de mots du dictionnaire le plus complet, en l'absence de néologismes. Cette limite ne devrait pas influer sur les textes de $B > 1$, pas plus que la vitesse de la lumière n'influe sur les phénomènes mécaniques ordinaires. Pour que l'influence de R_0 soit négligeable dans des textes usuels, il suffit d'une valeur modérée de R_0, mais pour les textes à information élevée (B proche de 1), il faut un R_0 très grand. Ceci explique la nécessité de la croissance du dictionnaire lors du développement des langues et projette un doute sur la possibilité de la construction d'une langue artificielle à R_0 limité (Basic English) qui ne présente pas de traits statistiques tout à fait pathologiques, comme c'est le cas en Basic English.

Etant données les difficultés conceptuelles et techniques qui s'attachent à toute grandeur "potentielle", le cas ouvert, qui n'en comporte pas, est conceptuellement et techniquement le plus commode.

Il est donc fort heureux que les courbes expérimentales se rangent dans leur très grande majorité dans cette catégorie et que la majorité des textes se caractérise par une seule variable d'état, le paramètre B.

Ce paramètre s'introduit dans la théorie tout-à-fait artificiellement, de la même façon que l'inverse de la température thermodynamique : comme "multiplicateur de Lagrange" d'un problème de calcul de variations. Nous appellerons par suite $\theta = 1/B$ la température informationnelle du texte.

B caractérise la variété dans l'utilisation d'un vocabulaire donné : B grand signifie utilisation de mots fortement concentrée sur les mots fréquents; B petit, utilisation plus largement répartie. C'est dans cette répartition plus large et l'incertitude plus grande sur le mot suivant qui en résulte que réside le concept de "quantité d'information" (qui est aussi complétement dissocié du "nombre de mots potentiel" ou de "richesse du vocabulaire").

Effectivement, la fonction d'état : quantité d'information, est une fonction de B qui décroît très vite (comme $(B-1)^{-1}$)

lorsque B croît (θ décroît) (Fig. 2, p.94). Malheureusement, B comme H sont des notions peu intuitives, de même que la longueur moyenne dans le codage optimum (qui toutefois varie dans le même sens que la longueur moyenne en orthographe réelle, donc est plus intuitive).

Il y aurait donc le plus grand intérêt à définir une "richesse de vocabulaire apparente" Q, variant en gros comme la richesse que l'on a tendance à attribuer instinctivement à un vocabulaire. On peut considérer que sont dans ce cas les "valeurs réduites" de R, qui auraient donné à l'une des fonctions d'état la valeur réellement observée, si B avait pris une valeur moyenne standard, par exemple 1.

En partant de H, on aurait $\frac{\log Q}{2} = H(B, \infty)$. En partant du paramètre P, qui est le plus facilement mesurable de tous, on aurait $P^{-1} = \log Q$. C'est cette valeur qu'adopte implicitement Shannon dans son étude de la prédiction de l'anglais écrit (1951). Il trouve ainsi Q = 8,727 mots.

Lors du vieillissement, on a l'impression que Q décroît, pourtant, en réalité, les mots rares ne disparaissent pas complètement, mais seulement deviennent "moins disponibles"; il serait donc intéressant de vérifier si le vocabulaire ne reste pas toujours équilibré, l'oubli se traduisant par une augmentation continue de B.

Réciproquement, la seule donnée disponible pour les enfants montre un vocabulaire équilibré et B décroissant de 5 à 6 ans. On peut estimer que la valeur minimum atteinte par B est une bonne représentation du plus haut niveau de verbalisation atteint par l'individu (si ce minimum est voisin de 1, il est très difficile à estimer exactement à cause des fluctuations) et la variation de B avec le temps, une mesure (à longue échéance) de l'efficacité des méthodes pédagogiques (qu'il serait, d'après ce qui précède, absurde de mesurer par le "nombre de mots connus" qui est une notion dépourvue de sens).

7.2.3 - INFLUENCE DE B SUR L'ESTIMATION DE R DANS LE CAS OU $B \leqslant 1$ (VOCABULAIRE FERME)

Dans ce cas, au contraire du précédent, la "somme d'état" P est divergente pour $R = \infty$. Une telle divergence est inconcevable en mécanique statistique des gaz et constitue l'une des originalités de la présente théorie.

Par suite, le nombre potentiel des mots différents R est nécessairement fini et toute révision d'une estimation de R a des répercussions considérables. En particulier, pour un texte très long, des nouveaux mots introduits pour réviser une valeur provisoire de R et rétablir l'équilibre devraient avoir, dès leur introduction, des fréquences comparables à celles des mots précédents. Donc ils n'auraient pas dû être omis dans le premier décompte. Par ailleurs, les mots nouveaux influençant beaucoup les fréquences des anciens, la notion de fréquence elle-même ne pourrait pas tendre vers celle des probabilités si $R \to \infty$ en même

temps que le nombre total de mots, car alors toutes les fréquences tendraient vers 0.

Pour R fini, la courbe donnant l'information en fonction de B a la forme de la fig.2 p.94. H vaut approximativement (log R)/2 pour B = 1, et log R pour B = 0 où H est maximum.

La variation de H (B) étant monotone, l'on peut remplacer la donnée de B et R par celle de $\theta = 1/B$ et H. Ceci sera pleinement analogue à la donnée de T et S en thermodynamique. Nous trouverons encore une autre description au § 7.2.4.

Le nombre des mots disponibles peut devenir important pour de multiples raisons. Par exemple, une langue ancienne, morte ou figée, peut être ranimée pour servir à un usage moderne, sans que le nombre de mots du dictionnaire soit augmenté par une règle automatique d'acclimatation des mots d'une langue étrangère, plus évoluée. On doit alors former des mots composés, en pleine conscience des limites du vocabulaire.

Ce fait entraîne le désir de transmettre une information supérieure au (log R)/2 du vocabulaire disponible, ce qui exige que l'on prenne $B < 1$.

C'est effectivement ce qui se produit dans l'exemple typique de l'Hébreu Moderne, où $B < 1$.

Un autre exemple est celui des mots latins de la "Mischprosa" de Notker : cette "prose mélangée" était un manuel de latin d'église à l'usage des moines. Il était en majeure partie écrit en gothique, de $B > 1$, mais les mots latins intercalés donnent $B < 1$.

Un dernier exemple est celui de certains poètes très "puristes" qui refusent d'utiliser la majorité des mots. Leur langue est un mauvais moyen de transmettre l'information. Elle est encore empirée par le fait que ces poètes se refusent à beaucoup d'expressions "fautives", ne pouvant de ce fait utiliser à plein les possibilités de la place des mots.

Un trait pathologique opposé, mais qui pourrait bien mener aussi à $B < 1$, est celui où l'on n'utilise pas les possibilités qu'offre le système des mots et utilise un nombre exagéré de mots spéciaux dans une langue peu riche. Ex.: Langue technique des marins, maquignons, etc...

Il serait intéressant à l'égard de la théorie B de comparer les valeurs que prend ce paramètre dans un texte authentiquement populaire et un texte "écrit en langue populaire". N'aura-t-on pas $B > 1$ dans le premier cas et $B < 1$ dans le deuxième ? Il serait également bon de vérifier si $B < 1$ dans le cas des aphasiques dont le vocabulaire est limité du fait de leur maladie.

Ce serait bien entendu un mérite pour la théorie que d'ouvrir ainsi divers problèmes nouveaux aux recherches expérimentales.

7.2.4 - PARAMETRES R ET Π

Pour introduire une nouvelle description d'un vocabulaire canonique, déterminé par le 1er critère, posons une décomposi-

tion de dC analogue à celle du dE de la Thermodynamique entre chaleur et travail.

$$dC = \theta dH - \theta p_R dR = \theta dH - \Pi dR = dL - dV$$

Les deux parties de dC s'interpréteront :

1°) dL comme variation de la "variété d'expression" à quantité d'expression constante (variable "noble"),

2°) dV comme variation de la quantité à variété constante.

P_R est la probabilité du mot le moins probable. π joue le rôle de variable d'état conjuguée de R. Si R est interprété comme volume, π doit être interprété comme pression.

L'"équation" d'état de (D) serait $\pi = \theta P_R\ (\theta,\ R)$. Elle prend pour $R \gg 1$, donc P_R très petit, la forme

$$R = (\theta - 1) \quad \text{pour} \quad \theta > 1$$
$$= 0 \quad \text{pour} \quad \theta < 1.$$

analogue à l'équation des gaz parfaits, qui est aussi une équation à faible pression.

7.2.5 - INTERPRETATION DE VARIATIONS DE H

A/ Si H croît, on a un processus d'apprentissage. Il ne peut se faire à B constant que si $R < \infty$. Alors R croît avec H, par exemple par mots nouvellement appris ou néologismes.

Si $R \doteq \infty$, la transformation d'apprentissage correspond à une variation de B, d'ailleurs très petite (car toutes les fonctions d'état varient très vite avec B voisin de 1), mais qui entraîne une grosse variation de Q (§ 7.2.2).

Au total, $H = \int \frac{dL}{\theta}$ depuis le début de l'apprentissage si on s'adapte à chaque apprentissage à la température θ correspondante. θ joue encore un rôle analogue à celui de la rempérature, C et H étant des fonctions des états initial et final, mais pas L ou V.

B/H peut aussi décroître : c'est par exemple le cas si on traduit un texte d'un vocabulaire étendu, de B voisin à 1, à un sous-vocabulaire, de B plus grand. Cette opération peut se faire sans secours extérieur si l'on utilise les définitions à la place des mots, mais le pourcentage des signes inutiles (redondance) croît dans l'opération. Comme décroît quand B croît, cette opération est, en un certain sens, l'équivalent de l'écoulement de la chaleur d'une source chaude à une source plus froide, qui peut également se faire sans secours extérieur. L'analogie entre et T va donc jusqu'au classement de ces quantités à l'aide d'opérations spontanées.

Une chute de θ est aussi en général une conséquence non souhaitée mais inévitable de toute traduction d'une langue dans une autre, car pour rendre la nuance juste d'un mot, on doit souvent le remplacer par une périphrase et très rarement le contraire.

7. 3 - PARAMÈTRES MACROSCOPIQUES DU CODAGE

Nous laisserons de côté m, et nous attacherons à l'étude de M, pour montrer d'abord comment il disparaît, (§ 7.3.2), ensuite comment on peut le faire reparaître, de façon expérimentale (§ 7.3.3), ou théorique (§ 7.3.4).

7.3.1 - LE PARAMETRE M

Dans l'étude de la signification physique de B qui vient d'être faite (§ 7.2), nous avons utilisé la formule simplifiée $p_n = P\ (n + m)^{-B}$. D'après cette formule, il n'y aurait pas deux mots de probabilité égale, celle-ci variant de façon continue du mot le plus probable jusqu'à des mots très rares. On a donc l'apparence de $M = 1 + \varepsilon$ pour toutes les courbes empiriques. Montrons cependant que cette apparence peut aussi s'expliquer avec M quelconque.

Plaçons-nous dans le cas simple, où M est le nombre de "lettres" équiprobables idéales utilisées dans le codage optimum. Deux mots ayant même nombre de lettres doivent être également probables. Par suite, si l'on tient compte de M (en négligeant m), la courbe p_n est, en coordonnées bilogarithmiques, une courbe en escalier dont les marches ont une hauteur et une largeur constantes et coupent la courbe approchée précédemment considérée. La largeur et la hauteur des marches croissent avec M. (La différence entre la courbe continue et les marches d'escalier correspond par exemple à la fonction périodique arbitraire g (log n) que J. Ville (1951) doit introduire dans une théorie (indépendante de la nôtre) où l'information d'un signal est donnée par la solution d'une équation fonctionnelle).

Nous appellerons k^e classe de mots de base M l'ensemble des mots faisant partie de la marche numéro k. k est le nombre de lettres du mot dans le codage idéal de base M. Il peut être pris comme mesure intrinsèque de la longueur du mot, c'est-à-dire en somme de sa "complication".

7.3.2 - DISPERSION

Sur un total de N mots, le mot M devrait apparaître approximativement p_n N fois. Cependant, ce ne peut être le nombre exact de sorties de M_n, car une courbe en escalier ne peut pas être une courbe de fréquence $f_n N$: $f_n N$ étant entier, la courbe de fréquence a des marches de hauteur constante en coordonnées linéaires. On passe de l'une à l'autre courbe en se rappelant que les fréquences correspondant à des mots équiprobables ne sont pas égales mais se répartissent de part et d'autre de la probabilité avec une dispersion donnée par la loi de Poisson répétée.

Cette dispersion émousse les différences entre marches voisines et donne, pour n modéré, l'illusion d'une courbe continue, c'est-à-dire de la courbe du § 6.4 et, pour n grand, d'une courbe en escalier à marches de largeur croissante.

Cette nécessité de considérer la dispersion de façon assez fondamentale est une différence profonde avec la Mécanique Sta-

tistique où les molécules sont si nombreuses que les fluctuations peuvent être traitées séparément comme termes correctifs. Ici elles font partie intégrante de la théorie.

Le seul cas où l'on puisse s'attendre à des effets de second ordre, attribuables aux marches, est celui où la dispersion maximum est très inférieure à la différence entre les probabilités de deux classes voisines, c'est-à-dire pour des échantillons très longs par rapport à l'information transmise par mots. (Effectivement, les courbes f d'un schizophrène font apparaître une cassure, de plus en plus marquée quand la longueur de l'échantillon croît). Dans tous les autres cas, la valeur de M ne pourra pas être déduite de la forme des marches.

7.3.3 - RECHERCHE EXPERIMENTALE DE M

On peut cependant mesurer M expérimentalement dès qu'on dispose de nombreux échantillons : En effet, tout en ayant peu d'influence sur la loi P_n, M influe sur l'ordre des mots, en ce sens qu'on peut affirmer qu'un groupe de mots est une classe dès que les membres de ce groupe sont sysmétatiquement plus ou moins fréquents que ceux des groupes voisins; tandis que l'ordre des fréquences à l'intérieur d'un même groupe varie entre deux échantillons pris tous les deux à l'intérieur d'un même texte homogène.

En prenant des groupes limités de plus en plus étroitement, par exemple en partant d'une valeur très grande de M, et étudiant ensuite des racines successives de cette valeur, on arrive à un moment où ce phénomène disparaît. Ce moment définit une valeur de M intrinsèque au texte.

La signification des classes de plus en plus étroites est bien entendu de plus en plus spéciale : les plus larges peuvent être considérées comme représentant de larges "couches culturelles" communes à des nombreux textes homogènes différents; les plus étroites ne s'appliquent qu'à des textes de moins en moins nombreux (c'est-à-dire des lignes de transmission de plus en plus particulières).

Par ailleurs, M étant donné, on obtient une classification des mots en classes différentes de la classification logique entre parties du discours grammatical.

Le problème de la détermination de M a donc une valeur intrinsèque. Il est d'un grand intérêt pour l'historien de la langue et de la littérature et c'est dans ce but qu'il a d'abord été étudié avec le même résultat : répartition des mots en classes. Plus précisément ce résultat était obtenu en cherchant à approcher une courbe expérimentale de fréquence par une courbe de Poisson composite. C'est ainsi que le statisticien Udny Yule (1944) a trouvé, en anglais moyen, trois classes en dehors des mots les plus fréquents, et un total "usuel" de 25.000 mots environ. Pour que ces classes soient en progression géométrique, il aurait fallu que M = 20 et elles auraient respectivement 50, 1000 et 20.000 mots. Cependant, le linguiste J. Guiraud a senti l'hétérogénéité des classes de Yule et il a été amené à décomposer chacune d'elles en deux sous-classes. Comme $20 = (4,5)^2$, les classes de Yule représentent la réunion en une seule classe de deux classes voisines de base 4,5.

En donnant successivement à M les valeurs 4 ou 5, nous devrons considérer respectivement 7 et 6 cycles différents ayant des nombres de mots respectivement égaux à :

4, 16, 64, 256, 1024, 4096, 16384

ou

5, 25, 125, 625, 3125, 15625.

7.3.4 - RECHERCHE THEORIQUE DE M

On peut aussi déterminer M théoriquement si, au lieu de se contenter de constater qu'il y a canonicité, on connaît le critère qui y a mené. Si on choisit une valeur moyenne de M et recode avec ce M des textes de M inconnu, les critères initiaux resteront approximativement satisfaits.

A) Critère probabiliste. Problème du secret.

L'existence de l'adaptation probabiliste entre la langue et le codage arithmétique signifie qu'il existe un code du type "à répertoire", c'est-à-dire codant le message mot par mot, tel que le message soit absolument indécodable pour qui ne connaît pas la clef authentique et parfait pour qui la connaît.

En effet, le problème du décryptage, tel que l'a présenté Shannon (1950), consiste à utiliser les propriétés statistiques du cryptogramme pour en déduire les propriétés structurelles du décodeur. Or, ici, le message crypté n'a aucune propriété statistique sur laquelle se baser. Il ne reste alors que la méthode exhaustive : essayer toutes les clefs; et même cette méthode nécessiterait l'hypothèse que l'ordre des mots par fréquence décroissante dans le texte est précisément un ordre standard, ce qui n'est vérifié que pour les mots les plus fréquents, qui sont les moins indispensables au sens.

Le fait qu'un tel code indéchiffrable existe est en opposition avec la supposition de Shannon, que les diverses conditions requises du code sont contradictoires. Ceci est dû aux propriétés statistiques du langage mot par mot, que Shannon n'avait pas considérées; il n'avait d'ailleurs pas à les considérer puisqu'il se plaçait dans le cas des codes lettre par lettre et non de codes "à répertoires".

Cependant, si le codage utilisé est "simple", les signes "top" peuvent être identifiés par leur fréquence (bien que ce soit peu utile) car il n'y a aucune raison pour qu'il y ait égalité entre la fréquence de signes ayant un sens et fréquence des tops. Toutes les deux dépendant à la fois du codage et de la longueur moyenne des mots, mais de façons différentes, qui font que ces deux fréquences sont en général différentes.

Cependant, si le critère probabiliste est satisfait, il existe une valeur de M et une seule qui rend les deux fréquences égales; donc les mots eux-mêmes impossibles à distinguer, si on n'a pas d'autres moyens d'identifier le top. Cette valeur est donnée par l'équation :

$$B = \frac{\log (M + 1)}{\log M}$$

Pour M =	2	3	4	5	6	7 ,
on a B =	1,58	1,26	1,188	1,113	1,086	1,069

Or, une bonne valeur moyenne de B est comprise entre 1,1 et 1,2, ce à quoi correspond M = 4 ou 5. D'ailleurs, lorsque B décroît et le vocabulaire s'enrichit, M croît aussi. Ceci est d'ailleurs extrêmement satisfaisant pour l'esprit, car ceci signifie que des codes nouveaux deviennent un peu partout disponibles pour de nouveaux mots, sans que l'on ait à modifier tous les autres codes.

B) Critère économique de prévoyance. Problème de la redondance.

Pour $C_0 = 1$, la relation de minimum de redondance : $C_0 = B^{-1} \log_M P$ devient une relation entre B, M et R. Pour $R = \infty$ elle s'écrit :

$$B = \log_M P = - \log_M (M^{(B-1)} - 1)$$

Pour les valeurs usuelles de M et de B, (B-1) log M est petit, et

$$B = \log_M (B-1) - \log_M \log_e M$$

Pour M = 5, B est encore compris entre 1,1 et 1,2.

7.3.5 - ORTHOGRAPHE REELLE

Admettons que M = 4,5 est une bonne valeur moyenne, et confrontons ceci avec la longueur moyenne des mots en Anglais.

Il se trouve que cette longueur est aussi 4,5. On pourrait donc recoder, dans un codage idéal, avec 4 ou 5 lettres sans allonger le texte par rapport à l'orthographe ordinaire. Celle-ci n'est donc en moyenne qu'une simple décompression résultant en l'emploi de trop de lettres, dans un désir de prononçabilité. Sa redondance serait : 1 - log 4,5/log 26 = 54 %, ce que confirment les expériences de Shannon (1951) qui sont indépendantes de nos raisonnements.

Remarquons que le dédoublement correspond au "principe de sécurité" des physiologistes : "les fonctions doivent pouvoir se poursuivre quand la moitié des éléments est détruite". Ce principe peut être compris de deux façons :

1°) Si les éléments sont les signes du message effectivement transmis, le dédoublement parfait doit assurer que le message reste compréhensible sans erreur dans le cas d'erreurs de transmission pas trop nombreuses (par ex. code auto-correcteur) ou qu'il devienne incompréhensible mais ne risque pas d'être confondu si les erreurs de transmission sont excessives.

2°) Si les éléments sont les mots du vocabulaire, le dédoublement parfait est tel qu'on puisse transmettre la même information quand la moitié des sources d'information primitives a été détruite.

Dans la réalité, une économie de moitié réalisée grâce au caractère séquentiel du codage (§ 6.4.3) est suivie d'une expansion uniforme qui double le message codé. Ces deux effets ensemble font que le principe reste satisfait dans ses deux acceptions ci-dessus.

CONCLUSION

Nous venons d'étudier de façon assez détaillée quel genre de rôle des modèles stratégiques sont susceptibles de jouer en Physique, au sens large, où ils se juxtaposent, à un niveau plus élevé d'abstraction, aux modèles mécaniques. Des uns comme des autres, on peut, avec plus ou moins de difficultés, déduire des propriétés susceptibles d'être vérifiées, et on peut leur appliquer le procédé d'induction physique.

Toutefois, tant que les précédents pour guider ce genre de théorie restent peu nombreux, il paraît souhaitable que celle-ci ne s'avance pas trop en pointe par rapport à l'expérience. On peut souhaiter que cette dernière oriente certaines études vers le développement des modèles tratégiques déjà proposés; après quoi, on pourrait entreprendre d'autres modèles, pour expliquer des nouveaux faits physiques, au sens large.

BIBLIOGRAPHIE

(Pour les revues, le nombre souligné est celui du tome)

BLUNDELL P.H. 1952 Applications of Communication Theory. Butterworths (sous presse)

BRIDGMANN P.W. 1943 Thermodynamics. Harvard University Press.

BRILLOUIN L. 1949 American Scientist 37, 554
1950 American Scientist 38, 594
1951 J. Appl. Phys. 22, 334 et 338

CALLEN H.B. and T.A. WELTON 1951 Phys. Rev. 83, 34

CASIMIR H.B.G. 1945 Rev. Mod. Phys. 17, 343

CHERRY C.E. et C.G. GOURIET 1952 Applications of Communication Theory. Butterworths (sous presse)

COURANT R. et K.O.FRIEDRICHS 1948 Supersonic Flow and Shock Wawes. Interscience, Publishers, New-York

DARMOIS G. 1935 C.R. Acad. Sc. Paris 200, 1265
1936 XXIIe Session de l'Institut International de Statistique Athènes

DEMERS P. 1944 Canadian J. Rescarch A22, 27
1945 Canadian J. Rescarch A23, 47

EHRENFEST P.& T. 1915 Encyclopédie des Sciences Mathématiques IV, 1, 6 188, Gauthier Villars, Paris

FANO R.M. 1949 M.I.T. Res. Lab. Electronics TR 65
1950 M.I.T. Res. Lab. Electronics TR 149

FISCHER R.A. 1922, 1935 dans (1950) Contribution to Mathematical Statistics. Wiley, New-York

FORTET R. 1950 Calcul des Probabilités. CNRS Paris

GABOR D. 1946 Journal I.E.E. 93 III, 429
1950 Phil. Mag. VII.41, 1161; ou La Cybernétique Ed. Revue d'Optique, Paris.
1951 Ritchie Lecture, Edimburgh

de GROOT S.R. 1951 Thermodynamics of Irreversible Processes North Holland Publishing Company

GUGGENHEIM E.A 1940 Thermodynamics. North Holland Publishing C°

GUIRAUD J. 1951 Thèse complémentaire. Faculté des Lettres de Paris

HAMMING 1951 Bell System Technical Journal 29, 147

HARTLEY R.V.L. 1928 Bell System Technical Journal 7, 535

HICK W.E. 1951 Research 4, 112

JAKOBSON R. 1951 Preliminaires to Speech Analysis. M.I.T. Lab of Acoustics TR 13

KHINTCHIN A.I. 1949 Statistical Mechanics (trad. Gamow) Dover

KOLMOGOROFF A. 1941 Bull. Acad. Sc. URSS Ser. Math. 5, 3

LANDE A. 1926 Handbuch der Physik. Theorien der Wärme

LAWSON J.L. and G.E.UHLENBECK 1950 Threshold Signals. Mac Graw Hill Book C°.

LEVY Paul 1948 Processus stochastiques et Mouvement Brownien. Gauthier-Villars, Paris

MACKAY D.E. 1950a Phil. Mag, VII. 41, 289.
1950b Communication Theory. Prof. W. Jackson. London

MANDELBROT B. 1951a C.R. Acad. Sc. Paris 232 1638
1951b C.R. Acad. Sc. Paris 232 2003
1952a C.R. Acad. Sc. Paris 234 1346
1952b C.R. Acad. Sc. Paris 234 1842
1952c Applications of Communication Theory. Butterworths (sous presse)

MAXWELL J.C. 1871 Theory of Heat

MIDDLETON D. 1952 Applications of Communication Theory. Butterworths (sous presse)

MOOERS C.N. 1950 Zator Technical Bulletins 30 et 31

MILNE E.A. 1948 Kinematic Relativity. Oxford U. Press

MOURIER E.A. 1946 C.R. Acad. Sc. Paris 223 712

von NEUMANN J. 1932 (et 1947) Fondements Mathématiques de la Mécanique Quantique, Springer (et P.U.F.)

von NEUMANN J.& O.MORGENSTERN 1944 Theory of Games and Economic Behaviour. Princeton University Press

NEYMAN J. 1937 L'estimation statistique traitée comme un problème classique de Probabilités. Hermann, Paris
1947 Procceding of the International Statistical Conferences III 423
1950 First Course in Probability and Statistics Henry Holt, New-York

NYQUIST H. 1924 Bell System Technical Journal 3, 324
1928 Phys. Rev. 32, 753

PAULI W. 1928 Sommerfeld Festchrift Leipzig

PRIGOGINE I. 1947 Etude thermodynamique des processus irréversibles. Dunod, Paris

PRIGOGINE I. et WIAME 1946 Experientia (Bâle) 2, 451

RASHEWSKY N. 1950 Bull. Math. Biophys. 12, 359

ROBERTSON H.P. 1949 Rev. Mod. Phys. 21, 378

de SAUSSURE F. 1916 Cours de Linguistique Générale. Payot, Paris
1949 "id" 4° édition.

SCHRÖDINGER E. 1944 What is Life. Cambridge University Press
1946 Statistical Thermodynamics. Cambridge U.P.

SCHUTZENBERGER M.P. 1951 C.R. Acad. Sc. Paris, 232, 925

SHANNON C.E. 1948 Bell System Technical Journal 27, 379 et 623
1949 Bell System Technical Journal 28, 656
1951 Bell System Technical Journal 30, 50

SHANNON C.E. and W. WEAVER 1949 The Mathematical Theory of Communication. Urbana, University of Illinois Press

SMOLUCHOWSKI M. von 1912 Physikalisches Zeitschrift, 13, 1069. 14, 261

SZILARD L. 1925 Zeits. f. Phys. 32, 653
1929 Zeits. f. Phys. 53, 840

TOLMAN R.C. 1938 Principles of Statistical Mechanics. University Press Oxford

UDNY YULE 1944 Statistical Study of Litterary Vocabulary Cambridge University Press

VILLE J. 1951 Câbles et transmission. 5, 76

VILLE J. et M.P. SCHUTZENBERGER 1951 C.R. Acad. Sc. Paris, 232, 206

WALD A. 1947 Sequential Analysis, Wiley, New-York
1950 Statistical Decision Functions, Wiley, N.Y.

WIENER N. 1948 Cybernetics, Hermann, Paris
1949 Time Series, Wiley, New-York

WOODWARD P.M. 1951 Theory of Observation. TRE Journal

ZIPF G.K. 1949 Human Behavior and the Principle of Least Effort. Addison-Wesley Press

TABLE DES MATIÈRES

INTRODUCTION

PREMIERE PARTIE

Chapitre 1

COMPORTEMENT INDUCTIF

Chapitre 2

INFORMATION ET DUREE

Chapitre 3

DEMONS DE MAXWELL

DEUXIEME PARTIE

Chapitre 4
THERMODYNAMIQUE DE SIGNAL PARFAIT

Chapitre 5
THERMODYNAMIQUE DU DECODEUR

Chapitre 6
CODAGE SEQUENTIEL

Chapitre 7
STRUCTURE STATISTIQUE DE LA LANGUE

REMARQUES
SUR LE
PROBLÈME DU CODAGE BINAIRE

par

M. P. SCHÜTZENBERGER

Hôpital Saint-Louis, Paris

INTRODUCTION

Soit le codage binaire optimum du point de vue de la fréquence totale des erreurs d'un ensemble de m messages de longueur l fixée à l'avance. Sa recherche pose un problème combinatoire pour lequel les méthodes de la théorie des communications sont d'assez peu d'efficacité. Le but de cette brève note est de montrer que ce problème se trouve déjà résolu en partie par les travaux des statisticiens qui ont été amenés à construire, pour un but tout différent d'ailleurs, des objets mathématiques, les "balanced incomplete block designs", dont il est possible de montrer qu'ils réalisent précisément ces codes optimum, tout au moins pour certaine valeur des paramètres.

DÉFINITIONS

Par définition, un code sera un ensemble de M messages $m_i = \{ X_i^J \}$ constitués chacun par une séquence de L symboles 0 ou 1. Il est naturel de supposer que des considérations de coût relatif imposent à priori une valeur déterminée au nombre N des symboles 1 dans le code et on posera :

$$N = \sum_i r_i = \sum_J k^J$$

où r_i et k^J désignent respectivement le nombre de symboles 1 dans le ième message et à la jème position sur l'ensemble de m messages.

Pour deux messages quelconques ou plus généralement pour deux séquences $\{y_a^J\}$ et $\{y_b^{*J}\}$ (on posera (cf. 6) :

$$2 D_{ab} = \sum_J \left[y_a^J (1-y_b^{*J}) + (1-y_a^J) y_b^{*J} \right]$$

D sera une distance entre $\{y_a^J\}$ et $\{y_b^J\}$ puisque $D_{ab} = D_{ba}$ et que $D_{ab} = 0$ entraine l'identité des deux séquences.

STRUCTURE DU BRUIT DE FOND ET DÉCODAGE

Nous supposerons que chaque symbole est transmis indépendamment et avec une probabilité constante p d'être reçu correctement.

Dans ces conditions il est évident que le vrai problème du décodage est exactement celui du choix entre plusieurs hypothèses tel qu'il est étudié en statistique mathématique.

Il est donc normal d'adopter une stratégie Mini Max consistant à interprèter la séquence reçue $y_a = \{y_a^J\}$ comme provenant de l'émission de celui des messages m_i tel que $Pr(y_a^J \mid m_i)$ soit maximum et en effectuant un tirage au sort avec des probabilités égales si plusieurs messages m_i se trouvaient vérifier cette condition.

Par conséquent, la discussion de l'optimalité d'un code devrait se faire sur la base d'une "information de Wald" c'est-à-dire en considérant des variables de la forme

$$Z_{ij} = \sum_a Pr(y_a) \ Log \frac{Pra(y_a \mid m_i)}{Pr(y_a \mid m_j)}$$

qui permettent cette discrimination entre les messages émis.

De fait, sous cette forme le problème semble inextricable et nous le remplacerons par le problème approché de trouver des codes tel que la valeur minimum de D_{ij} sur l'ensemble de toutes les paires de messages soit la plus grande possible. Nous appelerons pour abréger ces codes "codes optimaux".

LES CODES OPTIMAUX COMME « BALANCED BLOCK DESIGNS »

Théorème : Pour des valeurs données de M, L et N, il existe un code optimal si l'on peut construire une M x L matrice formée de 0 et de 1 telle que :

1°) Ses vecteurs lignes ont tous la même longueur,

2°) Ses vecteurs colonnes ont tous la même longueur,

3°) Le produit scalaire de deux secteurs lignes a une valeur constante.

Démonstration :

Calculons la variance de k^J sur l'ensemble des L positions

$$\text{var} \quad (k^J) = \sum_J (k^J)^2 - \frac{1}{l} (\Sigma k^J)^2$$
$$= \sum_J (\sum_i x_i^J)^2 - l^{-1} \quad N^2$$

par développement de $(\sum_i x_i^J)^2$ et permutations des deux sommations il vient :

$$\text{var} (k^J) = 2 L^{-1} (M^2 \quad L^2 - \quad N^2 - (ML \quad - \quad N)^2 \quad) - 2 \sum_{i,i'} D_{ii'}$$

comme var (k^J) est nécessairement non négatif, la valeur maxima de $\sum_{i,i'} D_{ii'}$ n'est atteinte que si tous les k^J sont égaux à une certaine valeur constante K.

Si cela est arithmétiquement possible, la valeur minimum de $D_{ii'}$ sur l'ensemble des couples de messages sera la plus grande quand tous les $D_{ii'}$ seront égaux. Mais pour un message m fixe, on a :

$$2 \sum_{ii'} D_{ii'} = \sum_{i \neq i'} \sum_J (x_i^J (1-x_{i'}^J) + (1-x_i^J) (x_{i'}^J) \quad) = r_i (M-K)+K (L-r_i)$$

u qui implique cette fois-ci $K \neq M - K$:

$r_i = R$ pour tout i.

Dans le cas où l'on aurait $K = M-K$, le raisonnement ne s'appliquerait pas, mais la conclusion reste sensiblement la même :

Prenons un message quelconque m_o. En permutant les symboles 0,1 entre eux, dans tous les messages pour certaines positions, on peut faire en sorte que $\{X_o^J\}$ soit toujours 1 sans affecter les $D_{ii'}$, et le raisonnement subsiste pour l'ensemble de ces m - 1 messages.

Les conditions énoncées par le théorème sont précisément celles qui définissent les "balanced incomplete block designs", tels qu'ils ont été introduits par F. Y. Yates pour les besoins de l'expérimentation statistique.

Sans entrer dans l'historique de cette théorie, nous rappelerons que le problème général de leur construction n'est pas résolu quoique l'on connaisse à la fois les solutions pour les faibles valeurs de M et de L (tables dans : (3)) et diverses méthodes plus ou moins générales de construction(Cf.en particulier (1) (8) et 10).

On sait d'autre part que les 5 paramètres sont liés par les deux relations diophantiennes :

$$L K = M R$$

et $\lambda (m - 1) = L \ (K - 1)$

où $\lambda = R - D$

et que l'on a toujours (R.A. Fisher 2)

$$L \geqslant M$$

Que ces conditions ne sont pas suffisantes a été montré pour la première fois dans (9) (et indépendamment peu après dans (12) pour une infinité de valeur des paramètres (si M = L = un nombre paire et si D n'est pas un carré parfait). D'autres résultats plus profonds ont été publiés, récemment (f.bibliographie dans (4) et (7)).

RÉFÉRENCES

1 Bose R. C. (1939) Ann. of Eug. IX 353-399

2 Fisher R.A. (1940) Ann. of Eug. X 52-75

3 " and F. Yates. Statistical tables for biological etc.. 3e edition. London 1948

4 Hall M. and Ryser H. J. (1951) Can. J. Math. (4) 495-502

5 Hamming R. W. (1950) Bell Syst. Tech. J. (26) 147-161

6 Laemmel A. E. (1952) Symp. on Appl. of comm.Theor.London

7 Mann. R. M. (1952) Can. J. Math. (4) 222-226

8 Rao. C. R. (1946) Proc. Nat. Ins. Ind. 123-135.

9 Schutzenberger M. P. (1949) Ann. Of Eug. XIV 286-287

10 " (1951) J. Roy. Stat.Soc. (B)XIII.120-125

11 " (1953) - Thèse à paraître.

12 Shrikande S. S. (1950) Ann. Math. Stat. (2) 106-111.

13 Yates F. (1936) Ann of Eug. VII 121-140.

J. & R. SENNAC, Imprimeurs
54, Fbg Montmartre - PARIS